ORD Scientific Information Management Implementation Coordination Plan 1999-2003

Revised Edition, February 10, 1999

Prepared for:

United States Environmental Protection Agency
Office of Research and Development
Science Information Management Coordination Board (SIMCorB)

Support provided by the
Technology Planning and Management Corporation Team

TPMC
Suite 310
4815 Emperor Boulevard
Durham, NC 27703-8427

TPMC
Suite 120
10306 Eaton Place
Fairfax, VA 22030

Suggested citation for this report is:

Shepanek Robert F., Gary J. Foley, Gary B. Collins, Linda L. Kirkland, Joan H. Novak, Jeffrey B. Frithsen, Michael Waters and collaborators (1999). Scientific Information Management Implementation Coordination Plan. Office of Research and Development, U.S. Environmental Protection Agency, Washington, D.C.

Message from the Chairman
ORD Science Information Coordination Board (SIMCorB)

I am pleased to provide you the ORD Scientific Information Management Implementation Coordination Plan.

Following its major reorganization in 1995 the Office of Research and Development embarked upon a concerted planning effort directed at realigning its goals, priorities and capabilities to meet the priority forecast science needs of EPA. The planning process has addressed three distinct levels of detail:

- *strategic*- definition of long-term, organization-wide missions, visions, goals and values.

- *tactical*- for each major organizational component (i.e., laboratories and centers) and specialized functional area (i.e., health, ecology, scientific information) articulation of program focus, management mechanisms, priorities and desired outcomes over the forthcoming three to five years intended to realize ORD's strategic vision.

- *operational*- articulation of discrete operational processes or projects to be implemented over a one to three year time frame, each designed to produce specific outcomes that contribute to realizing both tactical and strategic objectives.

The ORD *strategic* vision is reflected in three key documents: *Strategic Plan for the Office of Research and Development* (May,1996), *1997 Update of ORD's Strategic Plan* (April,1997), and the *Information Management Component* to the ORD Strategic Plan (August,1997).

This document presents the *tactical* plan for *scientific information management* across ORD for the five year period 1999-2003. In addition to definition of intermediate-level goals and objectives, management structures and organizational values, the document provides summaries of major projects to be implemented over the next one to three years. Because it is a blueprint to address issues, projects, and contingencies in the rapidly evolving disciplines of environmental science and information technology, it will undergo continuous refinement. At this stage, it provides a tactical platform from which ORD can move forward to engage science information management at the operational level.

Operational level project plans reflecting detailed descriptions, staffing planned outcomes, coordination processes, and resource requirements will be developed through SIMCorB interaction with the ORD laboratories and centers. These operational plans will be integrated into a coordinated business plan to support scientific information management throughout ORD.

This edition of the Plan incorporates consideration of issues raised in the SIMCorB Executive Advisory Committee peer review and the reviews conducted by the ORD Science, Management and Human Resources Councils. The Plan will be circulated through the CENDI[1] consortium to other Federal agencies with continuing interest in scientific information management.

Let me thank you in advance for your review of this Plan, and your continued support as we move from the strategic and tactical phases to the operational phase of meeting the scientific information management challenges that confront ORD.

Dr. Gary J. Foley

[1]CENDI, a consortium of Federal agencies concerned with scientific information management: Commerce, Energy, NASA, Defense, Interior.

Acknowledgments

The following individuals provided critical policy leadership and technical guidance in the development of this Plan:

Dr. Gary J. Foley, Chair, Science Information Management Coordination Board (SIMCorB)
(Director, National Exposure Research Laboratory)
Dr. Robert Shepanek, Vice Chair, SIMCorB
(Senior Information Scientist, National Center for Environmental Assessment)
Dr. Gary Collins, Executive Secretary, SIMCorB; Requirements Definition and Planning Sub-Group Chair (National Exposure Research Laboratory)
Dr. Linda Kirkland, SIMCorB Data Administration and Quality Assurance Sub-Group Chair
(National Center for Environmental Research and Quality Assurance)
Ms. Joan Novak, SIMCorB Advanced Technology Evaluation and Modeling Sub-Group Chair
(National Exposure Research Laboratory)
Dr. Jeffery Frithsen, SIMCorB Data Administration and Quality Assurance Sub-Group
(National Center for Environmental Assessment)
Dr. Michael Waters, SIMCorB Outreach and Liaison Sub-Group Chair
(Associate Director, National Health and Environmental Effects Research Laboratory)
Mr. John Ireland, SIMCorB Systems Engineering and Operations Sub-Group Co-Chair
(National Risk Management Research Laboratory)
Mr. Allan Sparks, SIMCorB Systems Engineering and Operations Sub-Group Co-Chair
(ORD Office of Resources Management and Administration)
Ms. Charisa Smith, SIMCorB Operations Sub-Group Chair (past)
(ORD Office of Resources Management and Administration)

SIMCorB Executive Advisory Committee: Donald G. Barnes (SAB); Wendy Cleland-Hamnett (OPPE-CEIS); John Convery (NRMRL); Mark Day (OIRM); Deborah Dietrich (ORD-CIO); Sidney Draggen (IOAA); William Farland (NCEA); Gary Foley (NERL); John Puzak (NCERQA); Richard Rineer (ORMA); Robert Shepanek (NCEA); Gilman Veith (NHEERL); Hal Zenick (NHEERL).

The following individuals associated with Technology Planning and Management Corporation (TPMC) contributed significantly to the development of this Plan: Jane Lovelace; JoAnn Kerrick; Thomas Hadd; Jolene Bollen; Anita Harris; Karin Forsythe; and Paul Celmer.

Table of Contents

Executive Summary

Scientific and technical information is central to all facets of the environmental protection effort - from discerning problems, to establishing understanding of environmental processes, establishing policy, evaluating risks, formulating preventative or remedial strategies, and evaluating results. Scientific and technical information takes many forms (e.g., reports, data sets, data bases, models, analytical tools, policy guidance, research plans) used by many different types of customers, including community-level problem solvers, research scientists and engineers, policy makers in Federal, State and local governments, and industrial executives and managers. The process by which scientific data and information is developed has become enormously complex. Scientific projects routinely involve multiple researchers, at multiple locations, and rely increasingly on sophisticated information processing capabilities to achieve their goals. The requirements of contemporary scientific studies demand a comprehensive system architecture that will support coordination of multiple paths of activity, facilitate analysis of distributed scientific information, provide efficient access to standard assessments tools, and assist in communicating results to a wide array of information users. The Office of Research and Development (ORD) has recognized the central role of scientific information management in maintaining the quality and integrity of scientific studies, and has established the Science Information Management Coordination Board (SIMCorB) to identify and address scientific information management (IM) requirements across the organization. The purpose of this document is to present a plan for ORD scientific information management and to outline the role SIMCorB will take in helping ensure this plan is implemented. The document moves from a high-level review of key components of ORD's strategic plan and concludes with specific scientific information management projects needed to fulfill ORD's strategic vision.

Current Status

No comprehensive scientific information management architecture exists within ORD. Separate applications, including the ORD Management Information System (OMIS), the Environmental Information Management System (EIMS), and domain-specific systems, are collectively used to track project finances and scientific data. This configuration has significant limitations; a single project cannot be seamlessly tracked from system to system, and system functionality cannot be shared. Information from multiple sources must be located separately, as no common point-of-entry exists to find and retrieve all scientific data. Tools including report writers, modeling and visualization applications, and geographic imaging software are implemented and maintained on a system-by-system basis. Assessments requiring information from multiple sources must be performed using third-party desktop applications. Public access is currently limited to those efforts using a web-based architecture and to those who have explicitly implemented Internet applications. In summary, collaborative work across the ORD

component organizations is severely impeded by a lack of coordinated scientific information management policy, procedures, and systems.

SIMCorB Strategy

The Environmental Protection Agency's (EPA) paramount goal is to protect human health and the environment: ORD SIMCorB exists to ensure that ORD's investments in scientific information management provide a solid scientific foundation furthering the pursuit of this goal. Major responsibilities include evaluating scientific IM policies, recommending strategic IM direction, providing leadership for multi-project initiatives, and sponsoring IM development projects. To meet these responsibilities, six standing Sub-Groups have been formed to address specific information management issues. These Sub-Groups currently include Data Administration and Quality Assurance, Requirements Definition and Planning, Systems Engineering and Operations, Advanced Technology Evaluation and Modeling, Science Direction, and Outreach and Liaison. Membership of each Sub-Group is comprised of interested parties with expertise in the appropriate technology or applicable project. The SIMCorB Executive Advisory Committee (SEAC) will provide continuing managerial guidance in the implementation process.

The activities coordinated by SIMCorB, SEAC, and Sub-Groups will reference a set of guiding principles to direct the management policies and values that determine cost, design, and operation of information management processes. Central to these principles is the commitment to leverage existing investments in information management through implementation of a distributed system: i.e., primary administrative, developmental, and operational responsibilities for existing systems will continue to reside with the sponsoring ORD Laboratory or Office.

A New Architecture

Addressing the current limitations to coordinated scientific information management, the conceptual design for ORD's Scientific Information Management System (SIMS) uses a web-based architecture to ensure ease of access for Agency and public users. Envisioned as a future version of the current EIMS, SIMS functions as a centralized locator for all ORD scientific information. The SIMS interface will allow users to find scientific data sets, databases, project descriptions, models, spatial coverages, and scientific and technical reports using a set of standard query criteria. The SIMS database stores metadata for each information object and allows actual data to be directly accessed, regardless of location, from the SIMS interface. Data residing in different databases may be analyzed using a standard online analytical processing (OLAP) tool and presented in graphical or report format. The environment provides universal access to information processing tools and standard third-party application software. Archived scientific information is readily available, and records management capabilities are enhanced. In addition, SIMS is able to transparently access and query existing ORD systems, including OMIS, to track a single project throughout its life

cycle. The environment facilitates information sharing and minimizes the possibility of duplicative efforts within ORD. Comprehensive management of ORD's scientific resources will enhance the quality and integrity of all information, as well as ensure usability for years to come.

Next Steps

The action plan focuses on the initiation and coordination of efforts that will define and implement major dimensions of the system architecture. Key policy and procedural issues will be defined and options developed for Agency action. In addition, cross-organization proof-of-concept pilot projects are planned. One such project will employ SIMS to integrate the National Health and Environmental Effects Research Laboratory (NHEERL) and the National Exposure Research Laboratory (NERL) endocrine disrupter research and link to external international information resources. This project will help to refine the roles and responsibilities of SIMCorB/SEAC members and the standing Sub-Groups. In addition, the project will serve to identify and refine standards for planning, data administration, quality assurance, operations, and systems engineering that will be used throughout the development effort to ensure compliance with accepted Agency procedures and compatibility between system components. Experience obtained throughout this development effort will be used as a model for future integration projects with SIMS.

Major emphasis will be placed on education and training about scientific IM concepts and tools. Outreach efforts will include identification of SIM user requirements and application capabilities, as well as definition of processes to permit two-way exchange of information with external scientific collaborators. The ORD Scientific Information Management Implementation Coordination Plan will continue to evolve as new requirements are identified and new projects are defined. Specific scientific information management projects will be periodically updated with status, activity, and outcome information.

Preface

These hypothetical cases illustrate how a coordinated Office of Research and Development (ORD) scientific information management (SIM) environment will function when this plan is implemented. The scenarios highlight leveraging the SIM environment to facilitate interdisciplinary, multi-investigator inquiry into research problems consistent with the strategic direction of ORD. One scenario illustrates a scientific study performed in response to a "pop-up" policy issue, the other illustrates planned science conducted according to the ORD research strategy. The scenarios will be referred to throughout the document to clarify and make more tangible proposed policy, procedural, hardware, software, and management approaches that will be employed to resolve the many scientific information management issues that currently face ORD.

The scenarios refer to planned modules of the SIM environment by their acronyms. A glossary of technical acronyms follows.

UV-B Case Scenario

A congressman from a southwestern state has been made aware by reports in the popular press of increased rates of skin cancer within his district. The story attributes the higher rates to changes in ultraviolet-B (UV-B) radiation. He requests the assistance of the Environmental Protection Agency (EPA) Administrator in determining if the increased rates of skin cancer really exist, and if they do, if UV-B is the cause. The Administrator tasks the EPA ORD with performing a rapid environmental assessment to provide sound and defensible information to assist the congressman in developing policy to protect the people of his district. Scientists within ORD had long been concerned with the implications of exposure to ultraviolet (UV) radiation for human and ecological health, and had devised a monitoring technology to measure potential exposure rates. A limited network of monitoring stations was established in collaboration with other EPA Offices and federal agencies to begin to better understand problems arising from UV exposure.

An ORD senior scientist within the National Center for Environmental Assessment (NCEA) is assigned the responsibility of coordinating the assessment project. Her first task is to identify individuals within the ORD laboratories and centers with expertise on exposure monitoring and environmental effects of UV-B on development of skin cancer. She turns to her personal computer and invokes the ORD Science Information Management Interface (SIMI). SIMI is the standard computer desktop for scientists in ORD. The ORD Science Information Management System (SIMS) is one of the selections available as part of SIMI. It provides her instant access to the ORD collection of data, information, tools, and metadata. Because the ORD collection is part of the "next generation" National Biological Information

Infrastructure (NBII) and the National Spatial Data Infrastructure (NSDI), her searches of the holdings of other governmental and non-governmental organizations are also facilitated using SIMS.

With SIMS she identifies scientists within ORD and at a university with the appropriate expertise. She requests that they participate as members of the research team. She directs them to an electronic discussion of the skin cancer/UV-B topic that she has set-up using Lotus Domino Internet groupware, another selection available through SIMI.

The result of the electronic discussion is consensus on the part of the team on how to frame the research questions posed by the congressman in a testable way. The initial hypothesis is stated, "skin cancer rates for the congressman's district are not increasing." The potential follow-on question is, "is the increased cancer rate due to increases in UV-B?"

The team uses SIMS to search for geographically and temporally relevant data, documents, and models that relate to UV-B, skin cancers, and human activity patterns. A variety of documents, data sets, databases, and models meet the criteria specified in this initial search of summary level descriptive information. The items are part of the holdings of the congressman's State Department of Environmental Protection (SDEP), the State University, EPA, the National Oceanic and Atmospheric Administration (NOAA), and the National Institute of Health (NIH). The team divides the work of evaluating each of the selected "objects" for use in this assessment based on the methods, detailed data descriptions, and quality assurance information accessed through SIMS.

Recent government agency reports and journal articles are linked to and reviewed online. The literature search indicates that cancer rates are increasing. Two of the selected data sets containing information on skin cancers are accessed using GEOSIM, the geographic information interface of SIMI. Because the data sets have different formats, the SIMS Data Format Wizard (SDFW) assists the team in setting-up access to them as if they were a single data set. The GEOSIM module is used to "clip" only data relevant to the congressional district. Time series analysis is performed on the "virtual" database using the statistical toolkit module to SIMI (STATSIM). The results of the analysis reinforces the conclusions from the literature search. Skin cancer rates in the congressman's district are indeed increasing.

The team uses SIMS to access the ORD UV-B database to retrieve summary information for instruments at Big Bend, Texas; Sequoia National Park, California; and Riverside, California. The team first scans the plotted distributions for a historical period spanning several years to see if there is an indication of increasing UV-B. Initial review indicates an increase. The SDFW is used to set up a virtual data set containing the relevant cancer and UV-B data. A spatial interpolation module in the modeling component of SIMI (MODSIM) is used to generate data within the congressman's district based on intensity data from several close

monitoring stations. STATSIM is employed to ascertain if a significant correlation between UV-B and increasing cancer rates exist.

After peer review, the information is summarized and integrated into a report for the congressman using the Internet groupware available through SIMI. Prior to delivery the draft is electronically routed through the ORD Assistant Administrator's Office to the EPA Administrator. The administrator's staff completes a search using SIMS for draft and pending congressional bills to identify any that might support mitigation efforts in the congressman's district. A bill is located that provides assistance to the states to increase awareness of health risks due to sun exposure. Information about the bill is included in the report for the congressman.

The final report is delivered to the congressmen along with a Web-enabled, multi-media presentation. The report is posted and available through SIMS on the ORD Intranet. Based on the report, the congressman supports the pending legislation, which passes.

Estuarine Research Scenario

An ORD scientist recently received an intramural grant award to conduct a study comparing nutrient loadings and phytoplankton production responses in several East coast estuaries. The scientist proposed and won this particular grant because of its relevance to multiple high-priority research issues outlined in the ORD Ecological Research Strategy (USEPA 1998) and its proposed implementation in the mid-Atlantic, one of ORD's priority geographic study areas. The initial filing of the grant proposal in ORD Management Information System (OMIS) starts the interaction with the SIM environment. To commence the project, the scientist compiles background information using the ORD Science Information Management System (SIMS), which provides an inventory of environmental resources collected or used by the Agency and its stakeholders and access to resource inventories in other scientific organizations. In addition, the SIMS link to the OMIS provides information on relevant current internal and external science activities sponsored by ORD. SIMS helps to identify several projects, data sets, documents, current researchers, and contacts focusing on estuarine nutrient enrichment and eutrophication. The ORD scientist uses PALMSIM to download this information into his palmtop laboratory and field notebook (PLFN) for more leisurely review that evening.

Using the information provided by SIMS, the investigator forms a research team comprised of Agency scientists representing ORD, the Office of Water, and Region 3. This team is brought together by video conference using the TELESIM module of the ORD Science Information Management Interface (SIMI). During the conference, the project leader outlines the project, discusses how the Agency data would be used to estimate point source nutrient loadings, and explains how estimates of nonpoint nutrient loadings would be derived from

various atmospheric deposition, runoff, and groundwater flow models. Additional modeling efforts to define the relationship between nutrient loading and phytoplankton production are identified.

The newly formed group uses the MODSIM module of SIMI to review what types of models might be used to generate loading estimates and to identify the types of input data needed to run those models. The team actively debates how to distribute resources between the chemists, who concentrate on estimating nutrient loading, and the biologists, who concentrate on phytoplankton response. Using the groupware meeting software provided through SIMI, the team quickly ranks these activities and assigns priorities. Data collection activities suggested by the biological modeling team surface as the highest priority.

Following the video conference, the project leader returns to SIMI, retrieves several relevant data sets for estuaries in the study region, and plots recently visited sample stations using the GEOSIM module of SIMI. Using STATSIM, a probability-based sample design is created to provide a sample density that will yield data with the precision and accuracy needed to develop the models discussed during the video conference. The researcher uses GEOSIM to plot the new sampling stations and locate the nearest boat launch ramps and marinas. Then, a field sampling plan is generated. The completed research project design is electronically transmitted to the OMIS project tracking system, and portions of the project design are extracted to create metadata describing the project.

During field sampling, crews track position using global positioning systems (GPS) that are linked to field personal data recorders (PDR). Both the GPS and PDR modules of PALMSIM are available on the scientist's PFLN. Latitude and longitude are recorded at each station, and station metadata are linked to the sample collection data recorded in the PDR. At the end of each field day, the contents of the PFLN (including sample metadata) are uploaded via satellite to the ORD SIMS, using PALMSIM. Samples are shipped to the appropriate laboratories.

EPA and registered contract laboratories use the LABSIM module of SIMI to learn about samples that will be received that day and, depending upon the analyses required, associate laboratory and analytical metadata with each sample. Following analysis, measurement data are entered through LABSIM. Thus, the various module of SIMI are used to link field and laboratory metadata with actual data.

Shortly after data collection and laboratory analysis activities are completed, the distributed research team use the STATSIM module of SIMI to review and visualize data, identify potential outliers, and compare new data with those that previously existed. A Lotus Notes discussion database linked to SIMS lets team members create updates concerning data verification and validation activities. Once the newly collected data has gone through most of the data maturation process, the SIMS data format wizard (SDFW) is used to create a database by merging previously existing data with new data. The chemical fate and behavior team then

uses MODSIM to develop estimates of point and nonpoint source nutrient loadings for each estuary of interest, while the biological effects team use MODSIM to develop estimates of phytoplankton production. Both groups use STATSIM to correlate model estimates with measurements. Model and analytical results are automatically saved by SIMS for subsequent review by the entire team.

Following all analytical activities, the team reassembles via video conference using TELESIM to discuss initial results. The team creates an outline for the presentation of project results and recommendations. Employing the document management (DOCSIM) and groupware capabilities of SIMI, the research team uses this outline to prepare a draft of their report. Results generated from the MODSIM and STATSIM modules and spatial data patterns from GEOSIM are brought together with the report text within the electronic document itself. A copy of the draft report is sent electronically to the appropriate ORD division director for review and comment. Following review, the electronic signature of the division director is placed on the report authorizing release of the report to the public. Updates of the project description are electronically recorded in the OMIS tracking system. Metadata describing the report are entered into the DOCSIM bibliography and, with the report itself, are published using the Agency Internet site the next day. All project documentation, data sets, models, and metadata are organized in electronic format awaiting management decisions on long term archiving.

Definitions of Terms

DOCSIM—Document management component of SIMI.

GEOSIM—Geographic information module of SIMI, a collection of GIS and visualization tools and routines for performing spatial analyses.

LABSIM—Laboratory information management component of SIMI.

MODSIM—Modeling component of SIMI, a suite of scientific modeling algorithms.

PALMSIM—Palm computer interface component of SIMS, used to link palm laboratory notebook and field data recorders to the SIMS collection.

PLFN—Palmtop laboratory and field notebook.

SDFW—SIMS Data Format Wizard; replaces the concept of a persistent common database structure with a knowledge-based front end that assists the user in finding and creating "virtual" databases as needed for the SIMS modules to access for analysis and reporting.

SIMI—Scientific Information Management Interface for the desktop, provides access to the library and collection of scientific data documentation and tools.

SIMS—Scientific Information Management System, a managed collection of scientific data, metadata, and documents shared by all scientists.

STATSIM—Statistical toolkit to SIMI, a comprehensive suite of statistical tools.

TELESIM—The teleconference component of SIMI.

1 Introduction

This section outlines the purpose, scope, and planning hierarchy of the Plan. It also defines the intended audiences as well as the intended benefits that users should receive once the Plan is implemented.

1.1 Purpose of the Implementation Coordination Plan 1999-2003

Scientific and technical information is central to every dimension of the environmental protection effort. Direct consumers of scientific data and information include policy makers at every level of government, industrial decision makers, those responsible for remedial or pollution prevention actions, and research scientists and engineers. Contemporary scientific activity can be enormously complex; projects may involve numerous research partners at multiple locations around the nation or the world. Smaller-scale science efforts rely increasingly on sophisticated information processing capabilities to achieve their goals. These characteristics of contemporary science impose exceedingly complex and frequently unique responsibilities on the information management (IM) function.

ORD has recognized the central role of information management in supporting the quality and integrity of science and has established an Information Management (IM) component of the ORD Strategic Plan. The ORD strategic IM vision highlights ORD's responsibility to provide defensible and credible scientific information in support of the Agency's mission and establishes the requirement that IM be included as an integral component of all ORD research planning and implementation.

The purpose of this Scientific Information Management (SIM) Implementation Coordination Plan is to define and create the IM support capability necessary to achieve ORD's strategic scientific goals. This plan is intended to be a "living document" and is subject to update as the array of current projects and cross-cutting initiatives are implemented, new information obtained, and new priorities established.

The logical sequence employed in the development of this implementation coordination plan is depicted below in Figure 1:

- Goals and vision derived from strategic documents and program plans
- Definition of the Science Information Management Coordination Board (SIMCorB) coordination and leadership roles
- Articulation of guiding principles used throughout SIMCorB activities
- System IM requirements derived from user and client needs
- Assessment of current IM organization and ORD laboratories and centers activities

- Articulation of an IM system vision/architecture
- Definition of specific implementation priority projects to link vision to operations

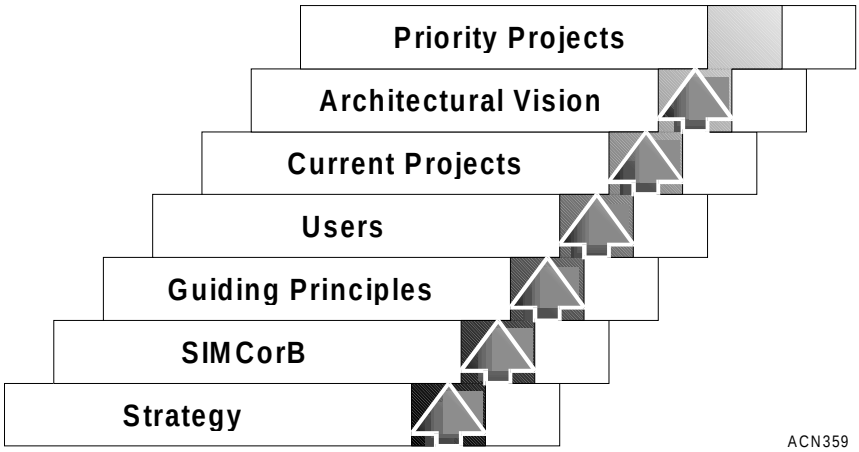


Figure 1. Implementation Planning Process

1.2 Perspective of the Implementation Coordination Plan

Scientific information management is one of several significant components of the overall ORD information management strategy. The many dimensions of scientific IM directly affect accomplishment of ORD's science goals and define the scope of this implementation coordination plan. These dimensions include a broad range of subject matter:

- Metadata concepts and standards
- Archival policies, standards, and technology
- IM standards for science project planning
- Scientific IM system architecture
- Data classification systems
- Data stewardship concepts and standards
- Communication processes and technology
- User education and training

The resources listed above permit the sharing of significant environmental information with scientific collaborators and stakeholders in environmental management efforts.

Some of the scientific IM requirements identified in this plan parallel needs of ORD-wide systems for management control and direction, internal and external communication, and data organization. Integrated strategies will be developed to meet these common needs. One notable example of such integration of effort is the proposed expansion of the ORD Management

Information System (OMIS) to include plans and documentation for all ORD-sponsored research (intramural and extramural) with particular attention to scientific IM issues. Similarly, ORD scientific IM actions must tie to the broader Agency-wide technology and IM structure if benefits are to be optimized. The Plan anticipates appropriate integration of scientific IM efforts with the overall EPA and ORD IM planning and priority-setting process.

A second major interface between scientific IM and ORD management processes is in the accountability and reporting structure established to meet the requirements of the Government Performance and Results Act (GPRA). The examples in Figure 2 illustrate the relationships among the fundamental structure of GPRA and scientific systems. The scientific IM strategy should include performance measures for scientific activities that link to the GPRA administrative reporting process. Such measures should be developed in conjunction with revisions to the research planning process.

The GPRA Planning Hierarchy: Outcomes—Examples

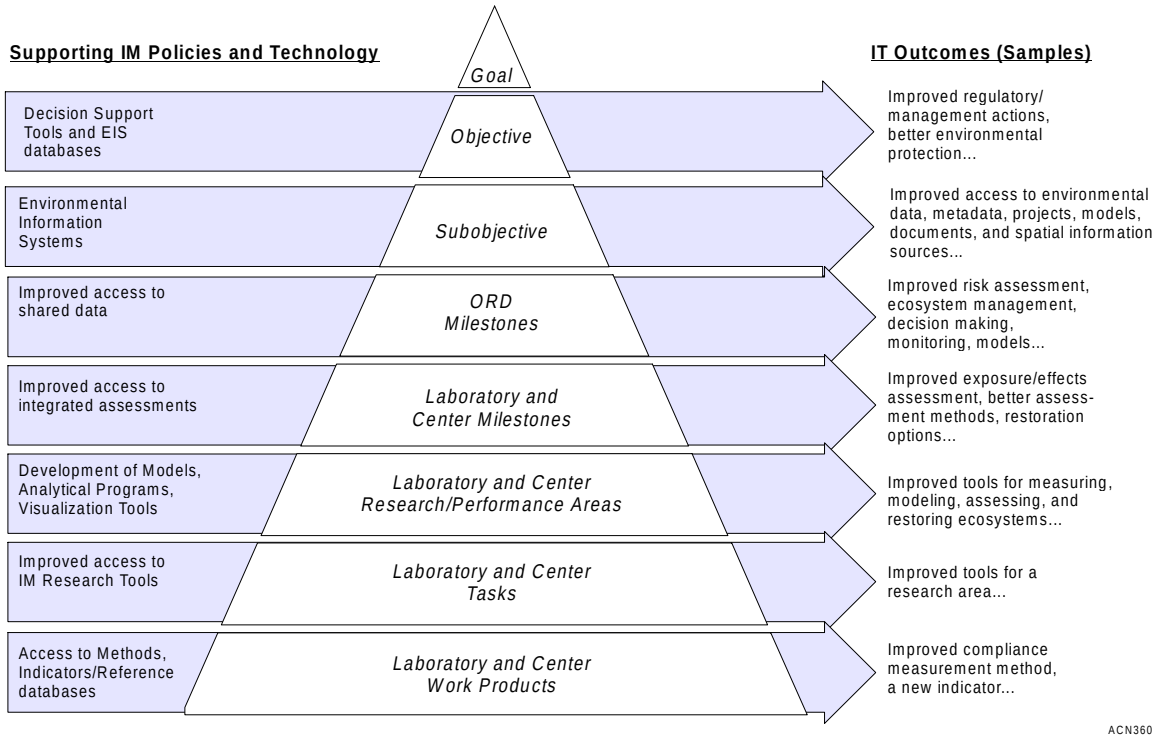


Figure 2. GPRA Outcome Pyramid

ORD scientific IM must also reflect requirements of the Information Technology Management Reform Act (ITMRA) which mandates that scientific decisions are made within a business context. The Plan should include procedures to perform system analyses and cost-benefit assessments complying with ITMRA specifications. A complete discussion of IM policy can be found in Section 2.2.2, *Organization and Policy for Information Management within ORD*.

This Plan outlines an approach for the development of policies, standards, and procedures, as well as technical information strategies, architectures, technical projects, and requirements definition for IM within ORD. The Plan addresses both cross-cutting issues and requirements, as well as the IM needs and projects of individual ORD organizations. Planned action projects scheduled for the next one to five years are outlined along with resource estimates, leadership responsibilities, and expected results.

1.3 Intended Audiences for the Implementation Coordination Plan

Information management issues touch every dimension of the science processes within ORD; thus, the relevant audiences for this plan are many, with each audience having varying requirements.

1.3.1 For ORD Scientists and Science Program Managers the Plan will:

- Provide a framework for planning IM dimensions of new projects.
- Create awareness of vision, goals, and IM resources.
- Create awareness of standards, policies, and emerging technology to serve ORD's needs.
- Create awareness of IM roles and opportunities for individuals to become active participants in development.

1.3.2 For External Scientists and Science Program Managers the Plan will:

- Create awareness of ORD IM vision, goals, and technology.
- Provide scientific IM framework and management structure to which they can relate their activities and programs.
- Identify specific activities that have relevance to EPA program offices or Regional Office operations.
- Convey SIMCorB's interest in establishing two-way scientific IM collaboration.

1.3.3 For External Consumers of ORD Scientific Information the Plan will:

- Provide a vision of planned pathways for public access to ORD scientific and technical information.
- Provide a framework and process by which requirements of these groups may be systematically identified and addressed.

1.3.4 For ORD Executives and Managers the Plan will:

- Provide a Scientific Information Management (SIM) blueprint to improve ability to meet EPA mission objectives.
- Provide an administrative vehicle for adopting a scientific IM vision and action program, and setting policy and investment priorities.
- Enable ORD scientists to cope with the complexities of global and integrated environmental research.
- Provide a mechanism for specifically identifying ORD's scientific IM requirements and linking these to the overall ORD IM strategy and action plan.

1.3.5 For the EPA Information Resources Management (IRM) Executive Steering Committee and Working Groups the Plan will:

- Provide an administrative vehicle to assure congruence and mutual support between EPA-wide IM efforts, and the requirements and strategy for ORD scientific IM.
- Identify specific areas where ORD will provide agency-wide leadership in the development and testing of cutting edge IM technology and processes.

1.3.6 For ORD and EPA Headquarters and Regional IRM Staffs the Plan will:

- Identify specific IM projects and priorities to facilitate coordinated implementation and evaluation.
- Minimize potential duplication of effort.

1.4 Intended Benefits Resulting from the Plan

The Plan establishes leadership in scientific IM within ORD through SIMCorB. SIMCorB, along with the SIMCorB Executive Advisory Committee (SEAC) and the Sub-Groups, will lead the researching, recommending, and implementing of a comprehensive scientific IM strategy for all of ORD. Furthermore, the Plan establishes a protocol for activities conducted by SIMCorB, SEAC, and the Sub-Groups by defining a set of guiding principles. Consistent application of these principles will result in increased effectiveness throughout all ORD scientific IM activities. Central to these principles is the commitment to support existing EPA

and ORD policy, including ITMRA and the Freedom of Information Act (FOIA), to leverage existing investments in research and technology, to promote data stewardship at all levels of the organization, to minimize duplication of effort, and to ensure the constant high quality of ORD scientific information.

The Plan and planning process provide a mechanism to achieve cooperative scientific IM within ORD and to share scientific and technical information outside of ORD in support of environmental protection. The hypothetical case scenarios presented in the Preface illustrated the intended benefits resulting form the Plan. From a communication and outreach perspective, the Plan outlines guiding principles and an IM strategy to achieve ORD's mission to provide sound scientific information in support of risk assessment and environmental policy. The Plan describes specific projects designed to allow ORD scientists and researchers to work collaboratively and to disseminate their results and conclusions to various external users. In addition, the Plan focuses on key information policy and procedural issues that, when resolved, will establish major operating parameters for the system development activities.

From a technological perspective, the Plan outlines an architecture designed to realize the functionality illustrated in the hypothetical case scenarios. The SIMS architecture specifies a centralized directory capability, enhanced Internet communication, and the capture and reporting of continuous monitoring data. These capabilities and others specified in the SIMS architecture allow data objects (e.g,. documents, data sets) residing at various locations to be easily found and accessed for assessment purposes, thereby promoting information sharing and cooperative research efforts among ORD laboratories and centers. This architecture also provides increased accessibility to environmental models, visualization, analysis and assessment tools, as well as the scientific information resources of other federal agencies. This suite of tools envisioned for SIMS allows users to perform assessment activities pertinent to their unique requirements without having the tool available on their desktop or local server.

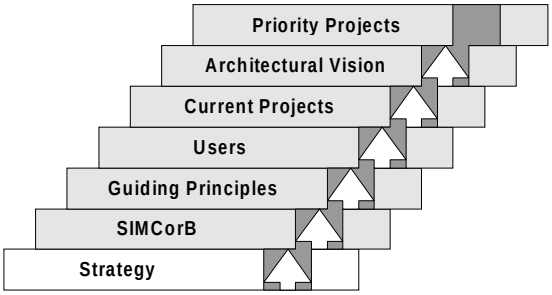
Primarily, the Plan prescribes a comprehensive assessment of ORD scientific information management requirements to meet ORD's mission and goals, and a procedure for ORD and EPA decisions on direction, priorities, and resource allocations to meet those requirements. Secondly, the Plan provides a technological road map to guide construction and acquisition of state-of-the-art IM hardware, software, and associated tools to support ORD risk management/risk assessment activities.

1.5 Organization of this Plan

Following this introductory section, Section 2 outlines the strategic policy framework that has been adopted to guide ORD information management efforts. Section 3 discusses the philosophy, mission, and organization of the ORD SIMCorB as the focal point for leadership on scientific IM issues. Section 4 highlights key principles that should guide specific design

decisions. Section 5 focuses on ORD scientific information users and specific categories of information requirements that must be addressed. Section 6 outlines the current significant science information management activities of the individual ORD operating organizations. Section 7 presents the architectural vision for ORD scientific IM. Section 8 summarizes the range of cross-cutting ORD IM project activity, resource requirements, and anticipated project results planned for next year and outlines dimensions of the IM vision for 2003.

2 ORD Strategic Vision



With the base provided in Section 1, this section maps out a high-level view of ORD's strategic vision and the importance of scientific information management in fulfilling that vision. The section also describes the planning process and the goals that will be met by the Plan. The IM component is explained through a discussion of planning, awareness, access, and usability. The organization and policy for information management in ORD is covered, and an introduction to SIMCorB is given at the conclusion.

2.1 ORD Strategic Planning Process—Strategic Plan

During 1995-1996 (and updated in 1997), ORD conducted an extensive strategic planning process that resulted in a comprehensive mission-vision-action strategy statement intended to guide actions across ORD for the next several years. Through this process, ORD defined its pro-active leadership responsibilities in achieving the following six long-term goals that lead from the Agency strategic guiding principle of strong science and data to support environmental decision making. Realization of each of these goals is directly linked to effective scientific information management. The six goals are:

1. Develop scientifically sound approaches to assessing and characterizing risks to human health and the environment.

2. Integrate human health and ecological assessment methods into a comprehensive multimedia assessment methodology.

3. Provide common sense and cost-effective approaches for preventing and managing risks.

4. Provide credible, state-of-the-art risk assessments, methods, models, and guidance.

5. Exchange reliable scientific, engineering, and risk assessment/risk management information among private and public stake holders.

6. Provide leadership and encourage others to participate in identifying emerging environmental issues, characterizing the risk associated with these issues, and developing ways of preventing or reducing these risks.

Collectively, these goals provide the policy framework within which the ORD Strategic Information Management Plan has been formulated.

The risk assessment/risk management paradigm that is incorporated in the ORD strategic goals has significant implications for scientific IM within ORD. Scientific and technical activities contribute to every stage of the risk management process. The flow and interrelationship of such activities is depicted in Figure 3 and captures the challenge confronting ORD in providing the scientific IM infrastructure and system support necessary to meet the long-term vision.

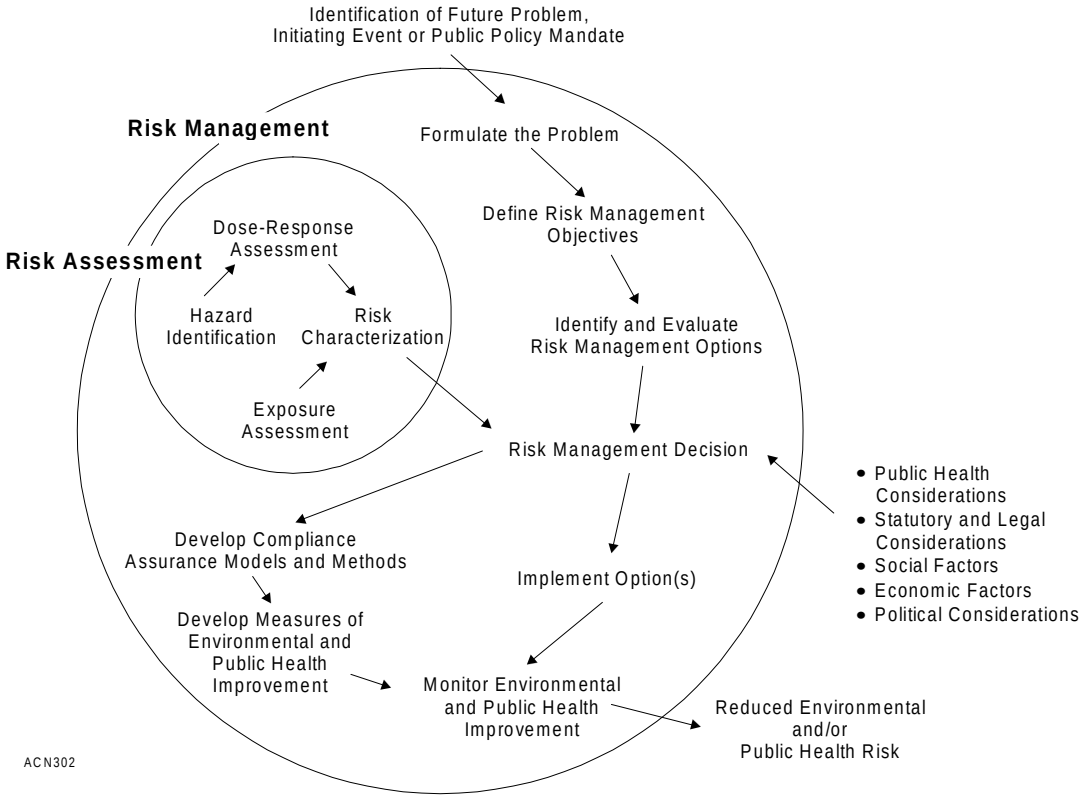


Figure 3. Scientific and Technical Contributions to Risk Assessment and Risk Management

2.2 Strategic Vision and Organization for ORD Information Management

2.2.1 Vision and Goals for ORD Information Management

The importance of information management to the accomplishment of ORD strategic objectives is clearly articulated in the Information Management Component to the ORD Strategic Plan (see Appendix, *Information Management Component*). That document presents, in detail, the programmatic rationale for the policy goals and operating objectives for scientific information management summarized below.

Four long-term goals for ORD IM to support risk assessment/risk management research were identified during the strategic planning process:

Planning

ORD will integrate IM planning into its research planning process to ensure that all of ORD's research information can be made available to potential users in a timely, effective, and efficient manner.

Awareness

ORD will provide the awareness, tools, and services needed to make its internal and external stakeholders aware of ORD's information.

Access

ORD will provide the communication paths and equipment that allow internal and external stakeholders access to ORD's information.

Usability

ORD will provide the planning, policies and standards, training, and user tools needed to make ORD's information usable to internal and external stakeholders.

These goals are congruent with, and supportive of, the Agency-wide strategic information management plan developed by the Office of Information Resources Management (OIRM).

With the focus on scientific IM, the four ORD goals translate operationally to the creation of a structure of technology, standards, and management processes to permit individuals or teams

of ORD researchers and their research partners, operating at a single site or geographically dispersed sites, to:

- efficiently define and document information requirements for the research activity.
- identify relevant scientific environmental information resources, held either internally or at a number of geographically dispersed locations.
- assess the utility and quality of those information resources for investigation of specific scientific issues relevant to environmental protection.
- gain access to the selected information resources.
- assemble, manipulate, and analyze the information resources developed through internal science, or acquired externally using state-of-the-art tools and technology.
- effectively present the results of such analyses to appropriate audiences.
- organize and archive data and information developed through ORD efforts to allow information sharing with other interested parties, and facilitate reproducibility of research consistent with the "20-year rule"[2] or other relevant record retention standard.
- acquire the knowledge, skills, and abilities to effectively employ ORD's IM resources.

The functional IM operating requirements associated with the environmental research process are the focus of this implementation coordination plan.

2.3 Organization and Policy for Information Management within ORD

ORD is organized into five major operating units and three headquarters policy offices that collectively conduct the affairs of ORD. Scientific information management affects every aspect of ORD operations and is a significant aspect of current ORD management and science processes. Following an in-depth review of scientific IM procedures, the ORD Assistant Administrator concluded that a new, comprehensive organization for scientific IM was not warranted.

A central tenet of the SIMCorB Implementation Coordination Plan is to utilize and enhance these existing management processes where possible to achieve more effective scientific information management. New roles and management processes will be added only to address voids that current arrangements cannot meet. Individual scientists have traditionally addressed scientific IM issues via the peer review and publication process. While this approach will continue to serve a portion of ORD's needs, scientific IM requirements imposed by the risk assessment/ risk management research processes require a significantly enhanced scientific IM capability to produce high quality, integrated scientific analyses supporting policy and action decisions.

[2] "Someone 20 years from now, not familiar with the data or how they were obtained, should be able to find the data of interest and then fully understand and use the data solely with the aid of the documentation archived with the data" (National Research Council, Committee on Geophysical Data, 1991).

A core value in ORD information management is that of "data stewardship." The concept includes recognition that much of the data and information generated in environmental management and science processes becomes a long-term resource which will be utilized in various ways long after it has fulfilled the initial purpose for which it was collected or generated. In addition to meeting Quality Assurance (QA) requirements planned for the original data use, the capability for secondary use entails planning requirements and implementation responsibilities for data documentation, adherence to applicable data standards, proper archiving, and accessibility to future users. This value has direct implications for everyone involved in ORD scientific data collection and use efforts, as well as those individuals with formal IM responsibilities.

2.3.1 Administrative Organization for ORD Information Management

Pursuant to ITMRA and EPA policy, the ORD Assistant Administrator has designated the Director of the Office of Resources Management and Administration (ORMA) as the ORD Chief Information Officer (CIO). CIO responsibilities include ensuring that IM activities and systems conform to ITMRA requirements. ITMRA legislation requires that scientific IM decisions are made in a business context and are supported by system analyses and cost-benefit assessments. The ORD CIO has been delegated authority to issue ORD-wide policies and procedures governing information technology and information resources.

Oversight and planning for administrative IM is centered in ORMA; administrative tasks are distributed between Automated Data Processing (ADP) Coordinators located at each ORD site. ORMA provides leadership in the design and operation of OMIS supported by an ORD-wide OMIS team.

The ORD QA management process addresses documentation and data integrity. Consistent with the strategy of using existing mechanisms for improving scientific IM, the 1998-1999 action plan calls for expanding the scope of the QA program to cover operating issues, including criteria and standards for management decisions on the archive and release of ORD scientific data and information. Overall, the ORD-wide network of QA Managers will become a central resource for realizing ORD scientific IM goals and objectives.

The technical information dissemination responsibilities of ORD also relate to the broader IM goals; the network of Technical Information Coordinators will play a key role in enhancing public access to ORD scientific information.

2.3.2 Scientific IM Roles and Responsibilities within ORD

As noted, the concept of "data stewardship" and associated values are central to the scientific ethic of ORD. Concern for the integrity and usability of ORD scientific information cuts across the entire organization, with specialized responsibilities at each organizational level.

Individual research scientists and engineers must assure that their research designs and scientific efforts address both immediate and long term requirements for utilization of scientific information. First line supervisors are responsible for assuring that SIM issues are fully addressed in all QA project and work plans, and that final products—both publications and electronic products such as data sets, models and model runs—are delivered and their records properly disposed of or retained. First line supervisors must also act to assure that appropriate SIM training and education opportunities are afforded to all staff. Managers are responsible for assuring that all components of their organizations are knowledgeable on SIM issues and concepts, that necessary technology and systems are available to support SIM requirements, and that organizational and individual performance and accountability standards are clearly linked to data stewardship.

2.3.3 Key ORD SIM Coordination and Policy Issues

Key policy issues that affect SIM within ORD are the following:

- Coordination—The ORD Strategic Plan noted the need for an ORD-wide coordinating focal point for scientific information management and called for the creation of SIMCorB, discussed below. In the process of developing this plan, members of SIMCorB identified the need for similar coordination efforts within each ORD laboratory and center. The action plan includes a focused effort to define needs, roles and responsibilities, and alternatives for establishing such a capability.

- Data Release—The ORD Science Council approved ORD policy on the release of research data in March 1998. The policy states that for intramural research, SIMCorB will define and refine issues and policies regarding "data keeping, retention, QA, format, etc. that apply to ORD supported research." Processes for designating data to be of continued research value versus "of such a routine, repetitive, or fragmentary nature that they would not be needed for continued research purposes" (Record Retention Schedule 503L) need to be clearly defined and integrated with planning for data storage, archiving, and disposal. For data of continued value, appropriate descriptive information in laboratory notebooks and/or project files is required.

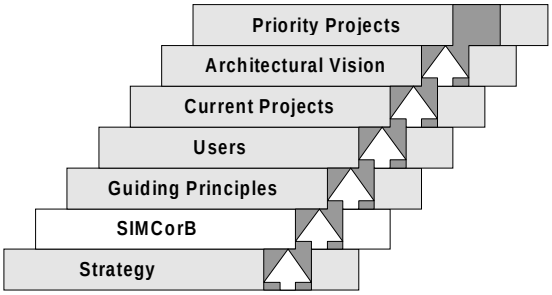
- Freedom of Information Act—ORD is responsible for responding to requests for scientific information under FOIA. ORD's final draft policy on research data states that upon completion of QA and peer review requirements, and either publication (paper or report) or use in an official government action, all data will be made available to the requesters. Access to research data (produced in-house or as a deliverable under an assistance agreement, interagency agreement, or contract) is appropriate because it was paid for by public funds, and its use in regulatory decisions can have impact on public health, the environment, the economy, and on individual private interests. Electronic access to scientific data for research carried out, or supported, by ORD is needed to minimize the

administrative burden of responding to FOIA requests while ensuring the continued integrity of data provided. SIMCorB will address search engine, security, and database design issues to establish usable access to stored or archived research data and its descriptive metadata.

2.3.4 Science Information Management Coordination Board (SIMCorB)

In response to the significance and uniqueness of many scientific IM issues, the high costs associated with information management, and the rapid pace of technological change, ORD established SIMCorB as a mechanism for managing scientific IM issues. SIMCorB is the principal advisor to the ORD Assistant Administrator and Executive Council on scientific information management issues, policies, and priorities. The detailed structure, membership, mission, and procedures of SIMCorB are described in Section 3 of this Plan.

3 Science Information Management Coordination Board (SIMCorB)



After the brief introduction and explanation in Section 2, this section goes into greater depth on SIMCorB by expanding on its charter, organization, Sub-Groups, and operations. SIMCorB's role in advising and supporting IM efforts throughout ORD is explained.

3.1 SIMCorB Charter and Organization

Established in 1997 by the Assistant Administrator for Research and Development, ORD SIMCorB serves as a permanent body for carrying out ORD's science information resources management responsibilities. The Board provides recommendations on key components of the ORD's science IRM program and policies to the ORD Executive Council, Management Council, and Science Council, and to the Assistant Administrator for Research and Development. The Assistant Administrator is also the formal ORD representative on the EPA Executive Steering Committee for Information Resources Management. SIMCorB is the ORD counterpart to the Executive Steering Committee for IRM. The ORD Chief Information Officer and the SIMCorB Chair provide operational representation of ORD to the Executive Steering Committee.

The purpose of SIMCorB is to ensure that ORD's investments in scientific information resources are managed efficiently and effectively in support of the ORD vision of providing the scientific foundation to support EPA's mission of protecting human health and the environment. In addition to sponsoring analyses and demonstration projects that will contribute to long term realization of strategic science IM goals and objectives, SIMCorB makes recommendations to the Assistant Administrator through the various ORD councils regarding approval of:

- ORD's Strategic IRM Plan, the Five Year IRM Implementation Plan, and associated budget issues.
- ORD scientific IRM investment portfolio.
- Management plans and concept papers for major scientific information system development, enhancement initiatives, and their respective budgets.
- ORD scientific IRM contracting strategy.
- ORD-wide scientific IRM policies and standards.

SIMCorB is chaired by the Director (National Exposure Research Laboratory [NERL]), who provides overall executive leadership and ensures ORD-wide perspective and cross-organization integration. The Senior Information Scientist (NCEA) serves as Vice Chairman and provides overall day-to-day program management and direction for the activities of the Board. The Laboratory Data Coordinator (NERL) is the executive secretary for SIMCorB and is responsible for logistical support for SIMCorB activities. Each major organizational component of ORD provides a representative to SIMCorB who understands the scientific mission of their organization and the principles of scientific information management. The representatives are empowered to make technical and resource decisions on behalf of the organization with respect to the management of scientific information. At the discretion of the Board, membership may be proffered to individuals outside of ORD.

3.1.1 SIMCorB Executive Advisory Committee (SEAC)

The ORD Scientific Information Management Implementation Coordination Plan, developed and maintained by SIMCorB, provides the road map by which ORD will attain its IRM vision. In the process of executing the implementation coordination plan, ORD will encounter numerous technical and management challenges. Whereas the Board is optimally configured to deal with implementing the technical aspects of the ORD science information management infrastructure, SEAC guides SIMCorB in overcoming management challenges to the successful accomplishment of ORD Strategic IRM goals and objectives. SEAC provides executive level input into the deliberative process of SIMCorB to assist in effecting change in long established ORD cultural and management paradigms regarding administration of scientific data, information, and tools. In order to develop positions and provide input to SIMCorB, SEAC will meet three times per year. In addition, members of SEAC will participate in SIMCorB functions and assist in defining and accomplishing the SIMCorB mission on as frequent a basis as they are able.

Representatives from the following organizations are proposed as members of SEAC:

- EPA Science Advisory Board (SAB)
- EPA OIRM
- EPA Office of Policy, Planning and Evaluation (OPPE)
- Office of the Assistant Administrator
- ORD—Chief Information Officer
- National Exposure Research Laboratory (NERL)
- National Center for Environmental Assessment (NCEA)
- National Health and Environmental Effects Research Laboratory (NHEERL)
- National Risk Management Research Laboratory (NRMRL)
- National Center for Environmental Research and Quality Assurance (NCERQA)
- ORMA
- Office of Science Policy (OSP)

3.2 SIMCorB Standing Sub-Groups

3.2.1 Standing Sub-Groups

The primary action component of SIMCorB consists of six standing Sub-Groups which are comprised of ORD staff from each of the ORD operating organizations. The Sub-Group responsibilities are generally oriented toward aspects of the system development life cycle, as follows:

- Requirements Definition and Planning (NCEA and NERL, co-leads)
    - Strategic and tactical science IRM planning
    - Gathering and documenting requirements for ORD science information planning
    - Development of plans to meet ORD science information requirements
    - Integration of requirements planning with other ORD processes
    - Translation of external user needs into technical system design requirements
- Data Administration and Quality Assurance (NCERQA lead)
    - Information and software security
    - Scientific metadata
    - Software quality assurance for applications and systems
    - Configuration and change management
    - Archiving of scientific information and data
- Systems Engineering and Operations (NRMRL and ORMA, co-leads)
    - Technology evaluation and selection
    - Application systems testing
    - Application system deployment and version control
    - Problem resolution
    - System level infrastructure
    - Access control
    - Application performance monitoring
- Advanced Technology Evaluation and Modeling (NERL lead)
    - Identification, evaluation, and testing of new technologies
    - Development of plans for integration of new technologies into the ORD mainstream

- Development of proof-of-concept pilot projects
- Evaluation and validation of scientific models

- Science Direction (NHEERL and NOAA co-leads)

  - Monitoring strategic directions in science that could provide opportunities to advance scientific information management.
  - Monitoring of IM initiatives of other scientific organizations and Federal agencies whose programs and activities are relevant to ORD scientific IM.
  - Acquisition of scientific data from external sources for systematic integration into ORD risk assessment/risk management activities.

- Outreach and Liaison

  - Organization and dissemination of ORD scientific data and information within EPA and to external audiences.
  - Training and education relevant to scientific IM for all users.
  - Fostering two-way communication between ORD scientists and engineers and external scientific information users.
  - Promoting ORD access to external scientific information resources.
  - Fostering access and usability of EPA/ORD scientific information by external audiences.

3.2.2 Sub-Group Organization, Management Information System (MIS) and Membership

Leadership of Sub-Group efforts are provided by chairpersons selected because of their unique technical and leadership capabilities. Each chairperson will be a member of SIMCorB. Sub-Group membership will be comprised of ORD staff and other partners with expertise and interest in the responsibilities of particular Sub-Groups. In general, chairpersons will devote 25% or more of their work effort toward SIMCorB responsibilities. Time requirements for individual members may vary based upon their respective project involvement.

The mission of each Sub-Group is to understand the subject matter within its scope of responsibility, survey needs and requirements, frame issues and recommendations for consideration by SIMCorB, identify and recommend to the Board IM development projects to be implemented by ORD line organizations, and assist in implementation of adopted IM policies and procedures.

Each work group will be responsible for developing and maintaining a body of policies, standards, and procedures for ORD science information management in the specific area of their responsibility. These policies, standards, and procedures, as well as changes to them, will

be subject to review by the full Board which is empowered to accept or reject them on behalf of ORD. The guiding principles presented in Section 4 are organized in accordance with SIMCorB Sub-Group responsibilities. The Sub-Groups are expected to monitor ORD compliance with the principles over time.

The membership of each standing Sub-Group will be the focal point within ORD for scientific IM leadership within their respective subject matter areas. As such, they will serve as a source of technical information and advice for all ORD operating units, individual scientific project or program managers, and to Agency-wide IM managers as appropriate.

3.2.3 Special Focus Ad hoc Sub-Groups

Ad hoc Sub-Groups may be established by SIMCorB to address special issues, such as the cross-cutting requirements for a fully developed IM system to support integrated risk assessment/risk management activities in a major science program area. The project manager will be responsible for identifying issues relevant to the charters of the individual SIMCorB Sub-Groups and establishing the necessary coordination.

3.3 SIMCorB Operations

SIMCorB provides a unique organization for blending technical and administrative perspectives on a series of extremely complex scientific information management issues. SIMCorB operations will focus on four key areas as described in the following sections.

3.3.1 Evaluation and Adoption of Scientific Information Management Policies

Because its composition provides for perspectives from management, as well as technical leaders for all major aspects of scientific IM, SIMCorB has the capacity to evaluate current policies and procedures impacting IM, identify problem areas, and develop and assess options for improvement. The major focus will be the ORD and Agency-wide policies and procedures necessary to achieve implementation of the overall Scientific IM architecture vision. Specific policy and procedural changes will be adopted by SIMCorB and recommended to the appropriate management organization for implementation.

3.3.2 Recommendations on Priorities and Strategic IM Direction

SIMCorB will periodically review the ORD strategic and science program planning documents and evaluate them in relation to emerging scientific information management issues. Of particular importance are the Five Year IRM Plan, the ORD scientific IRM investment portfolio, new scientific program starts identified in the annual planning process, and the

structure of IM investments incorporated in the annual budget proposals. The Board and Advisory Committee will make recommendations to the Assistant Administrator on priorities and strategic direction for IM investments.

3.3.3 Coordination and Leadership on Cross-Cutting Issues and Multi-Project Initiatives

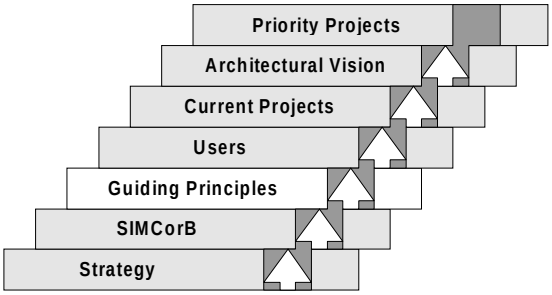
SIMCorB is uniquely positioned to identify IM issues and opportunities that affect two or more ORD organizations or science programs, and to initiate coordinated actions to establish mutually beneficial solutions. Similarly, the Board has the breadth of vision and technical competence to shape broadly focused IM initiatives (e.g. metadata documentation standards and techniques) and coordinate multiple projects that support accomplishment of the vision. The proactive leadership responsibilities of the Board and Advisory Committee in these areas are central to realization of the ORD strategic IM vision.

3.3.4 Sponsorship, Oversight and Management of IM Development Projects

The majority of IM development projects will be initiated and implemented by operational ORD organizations following specific program directions. However, opportunities for efficiency and cost-effectiveness in meeting new needs, adapting relevant technology, or addressing voids in the IM system architecture may require SIMCorB to define and sponsor IM development projects intended to have future, cross-cutting impact.

For each SIMCorB development project a senior EPA manager or scientist working within one of the ORD centers will be designated as Project Manager to provide management direction to the activity. This will include preparation of project plans for SIMCorB review and approval, identification and mobilization of staff and financial sources, liaison across ORD or EPA as appropriate, and periodic reporting to SIMCorB or ORD management as required. The Project Manager will be responsible for identifying issues relevant to the charters of the individual SIMCorB Sub-Groups and establishing the necessary coordination.

4 Guiding Principles for ORD Scientific Information Management



This section outlines the specific guiding principles for IM, policy, data and data administration, system design, operations, technology, and outreach and liaison in reference to ORD scientific information management. In addition, these principles will be used to guide the development of the cooperative scientific information management environment illustrated in the hypothetical case scenarios. This section marks the scope of SIMCorB's activities. SIMCorB has a central role in helping ensure these guiding principles are taken into account during the planning and execution of ORD scientific research and development.

4.1 Application of Guiding Principles

The guiding principles presented in this section are a formal restatement of values, rules, standards and codes of conduct intended to govern the conduct of ORD activities as they relate to scientific information management. These guiding principles set boundaries for appropriate behavior, set standards for performance evaluation, and provide guidelines for decision making.

The development of an organization-wide scientific information management capability requires the participation of many individuals at various locations, each bringing potentially different assumptions to guide their efforts. Consistent application of these guiding principles should minimize disruption and ease coordination requirements among and within ORD laboratories and centers.

These principles also carry implications for the cost, design, and operation of information management processes. For example, the principle that scientific information derived from certain categories of research should meet the "20 year rule" has implications for archive and release policies, archive technology, communications technology, and so forth. Thus, the guiding principles are not just a neutral road map, but also a statement of management policies and values with clear operational implications.

4.2 Guiding Principles on Science and Information in Environmental Protection

The following guiding principles establish the policy context for scientific information management in ORD:

- EPA has identified strong science and credible data as one of the guiding principles to fulfill the Agency's mission to protect human health and environmental quality.

- ORD has leadership responsibility for science at EPA, recognizing that other EPA Offices contribute to the scientific underpinnings of the Agency's decisions; ORD will provide leadership in establishing information standards and in research on computer technology to enhance the production and flow of scientific information.

- Effective environmental protection demands an information-rich setting in which all stakeholders—including policy makers, industry and the regulated sectors, and community participants—have effective access to the full array of useable environmental information and data needed for forming judgements on conditions and decision making.

- Access to scientific and technical data and information generated through publicly funded research or development projects should not be restricted; the timing and method of full access to such data and information should be explicitly stated as a condition of the provision of public financing.

4.3 Guiding Principles for ORD Scientific Information Management

ORD has adopted the following guiding principles for Information Management:

- The quality of ORD scientific information should be known and understandable.
- Information systems and documentation standards to demonstrate quality must be established and their integrity maintained.
- Achievement of scientific information goals and adherence to IM standards is the responsibility of all scientists and program managers; the methods and costs for meeting scientific IM requirements will be explicitly recognized and documented in the design of each ORD scientific project.
- Scientific data and information (including project design documentation and funding information) resulting from publicly funded ORD scientific activities should be made publicly available in a timely manner and in user-friendly formats.

4.4 Guiding Principles for Policy

ORD has adopted the following guiding principles for policy issues:

- ORD will be a catalyst for information standards.
- For IM systems to succeed, they must be built upon information standards. ORD will actively identify, prototype, and influence information standards internally, within EPA, and across other agencies and organizations that produce such standards.
- ORD will adopt and implement pre-existing Agency IM policies whenever possible.
- ORD scientific IM will examine existing EPA, Federal, and scientific information management policies and incorporate existing procedures that fulfill ORD system and user needs.
- Collection, analysis, and management of scientific data and information in support of risk assessment/risk management is complex and costly. ORD IM must focus on the efficient, cost-effective deployment of appropriate information technology and associated policies and procedures.
- ORD will provide leadership in the identification and resolution of policy issues that impede realization of the scientific information management vision reflected in this Plan.
- ORD will provide leadership in restructuring rewards and incentives to promote full acceptance and adoption of the scientific information management values reflected in this Plan.

4.5 Guiding Principles for Data and Data Administration

ORD has adopted the following guiding principles for Data and Data Administration:

A core value in ORD data administration is data stewardship; all individuals involved in data generation, manipulation, reporting and/or storage have data stewardship responsibilities to ensure that their actions protect and enhance the long-term usability of the information under their control, including adherence to adopted data standards.

- ORD scientific information is comprised of data and metadata.
- Metadata is descriptive information about data that explains the reason for the data collection, or generation through models; the methods employed in collection/generation and analysis of that data; and the procedures used to ensure its integrity. Data is logically and functionally inseparable from metadata. Raw data has little value without metadata descriptions (e.g., sampling location, sampling techniques, analysis methods, and quality control procedures). Data should not be distributed or reported without its associated metadata.
- Scientific efforts that rely on multiple and/or large, separately collected datasets have added information management requirements.

- All scientific projects sponsored by ORD are subject to documentation, quality control, archival, and directory recording policies adopted by ORD. These standards, and those associated with publication in referenced scientific journals, will normally provide metadata requisites for replication of the scientific activity and/or independent verification of conclusions. Those ORD scientific efforts that mandate the use and manipulation of large datasets drawn from diverse sources should adhere to policies and procedures that will permit independent assessment and replication of the analysis, ensure that data quality has not been compromised through the analysis, and ensure that complete metadata is available for potential secondary users of the data and associated analyses. Documentation and archival requirements will be established by ORD management for each sponsored research project prior to initiation of the project.

  A key standard for metadata documentation is the "20-year test." "Someone 20 years from now, not familiar with the data or how they were obtained, should be able to find the data of interest and then fully understand and use the data solely with the aid of the documentation archived with the data" (National Research Council, Committee on Geophysical Data, 1991).

- Information integrity should be managed. Data and metadata integrity should be maintained. This includes a common definition and tracking of information pedigree, archiving, quality, and physical security (including backup and recovery).
- Information integration and distribution require standards. Data conventions and standards should be developed and adhered to by all participants. The development of such standards requires close collaboration between IM specialists and scientists, engineers, and managers. Standards should include elements such as metadata codification, information exchange protocols, and translation techniques.

4.6 Guiding Principles for System Design

ORD has adopted the following guiding principles for system design:

- Economy, efficiency, and interoperability over time should be key criteria in system development.
- IM systems should be designed to support the risk management/risk assessment mission of ORD.
- System designs should employ a common language to describe the function and operation of system components.

4.7 Guiding Principles for Operations

ORD has adopted the following guiding principles for operations:

- Operations should provide secure, quality service and regular review, with focused effort on meeting user needs.
- Operations should facilitate sharing of data and data manipulation software.
- Operations should enable effective communication of project and scientific information.

4.8 Guiding Principles for Technology

ORD has adopted the following guiding principles for technology:

- Technology will constantly change.
- ORD IM should anticipate and accommodate the eventual changes that will occur in information systems technology, scientific analysis techniques, communication process, and associated information handling technology.
- Risk assessment/risk management requires access to state-of-the-art information processing technologies.
- The complex functional requirements created by integrated risk assessment/risk management require a focused effort for seeking out and evaluating emerging new technologies for potential application to ORD mission requirements.

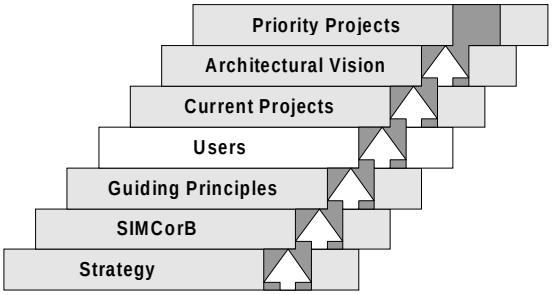
4.9 Guiding Principles for Outreach and Liaison

ORD has adopted the following guiding principles for outreach and liaison:

- ORD scientific data and information should be made available to the full array of interested external customer groups—including industry, community participants, policy makers, and action officials—in a timely manner and usable formats.
- ORD will collaborate with the EPA Center for Environmental Information and Statistics (CEIS) in discerning the scientific and technical information requirements and accessability issues of various customer groups and develop effective responses.
- Communication processes for ORD scientific data and information should be structured to allow user definition of requirements.

- ORD, in cooperation with the Office of Information Resources Management and the Center for Environmental Information and Statistics, should systematically investigate and understand the scientific IM direction of other science organizations and other Federal agencies whose programs link to ORD risk assessment/risk management responsibilities and promote collaboration where appropriate.
- Training and education relevant to scientific IM should be provided for ORD scientists and managers, as well as for the general public.

5 User Requirements for ORD Scientific Information



Each group of users has a unique set of scientific IM requirements. This section provides a summary of a selected group of internal and external users, outlines their respective information management backgrounds, and identifies specific requirements for each set of users. In addition, this section demonstrates how these user requirements will be satisfied by the Plan, and how SIMCorB's assistance can benefit each type of information user.

5.1 ORD Scientific Information

5.1.1 Users of ORD Scientific Information

The full range of potential ultimate users of ORD scientific information is difficult to forecast with precision. Environmental and public health assessments, whether conducted by scientists or others, inherently rely upon scientific information. Thus, the potential audience for ORD scientific information is immense. The EPA CEIS is sponsoring a comprehensive survey of customer requirements for environmental data and information that addresses the general interests of the following major constituencies:

- neighborhood and community environmental organizations
- the agricultural and mining communities
- State environmental departments
- Native American and Ethnic community leaders
- environmental researchers and environmental quality activists
- manufacturing and business interests
- public health organizations

Preliminary analysis of the results of this survey demonstrates the significant diversity of information and communication requirements among external environmental information users that must be considered in the design of a comprehensive outreach strategy. Recognizing that the technical needs of many groups will require further documentation, the results of the CEIS efforts will be one key resource for the Outreach and Liaison Sub-Group as it develops recommendations for SIMCorB consideration on the most effective means for addressing the priority needs of the diverse constituencies for scientific and technical environmental information.

Recognizing this broader universe of potential users, this Plan highlights the scientific information requirements of five major users groups and identifies how these needs will be addressed over time:

- ORD scientists, engineers and risk assessors
- ORD executives and managers
- policy makers and decision officials in government and industry
- stakeholders in community-based environmental problem solving
- external scientists

Each of these groups has different requirements for data resolution, classification, and focus. Effective scientific information management requires that the specific needs of all relevant stakeholders be identified in a systematic manner in relation to concrete scientific activities. Consequently, the requirements analysis process for each management function, program, or area of scientific activity should identify the type of information required, the purpose, the degree of resolution, and how the needs may be efficiently addressed.

5.1.2 Environmental Science in ORD

The environmental science process of ORD, illustrated in Figure 4 below, highlights the types of activities performed by ORD laboratories and centers during generation and assessment activities. Once a hypothesis has been formulated, experimental design activities commence. Data is collected, and data models are generated. The final step, wherein the data is analyzed and assessments are formed, leads to the re-formulation of the hypothesis and repetition of the process; if the initial hypothesis is correct, the project is considered complete, and the cycle ends. As illustrated, the cycle depicts both assessment activities (data collection and assessment formulation) and generation activities (data modeling and analysis) that allow the environment to fully support the various activities of NERL, NHEERL, NCEA, and NCERQA. Successful scientific information management within ORD logically requires that the information environment fully support each procedure performed by ORD laboratories and centers during the environmental science process and communication with external audiences.

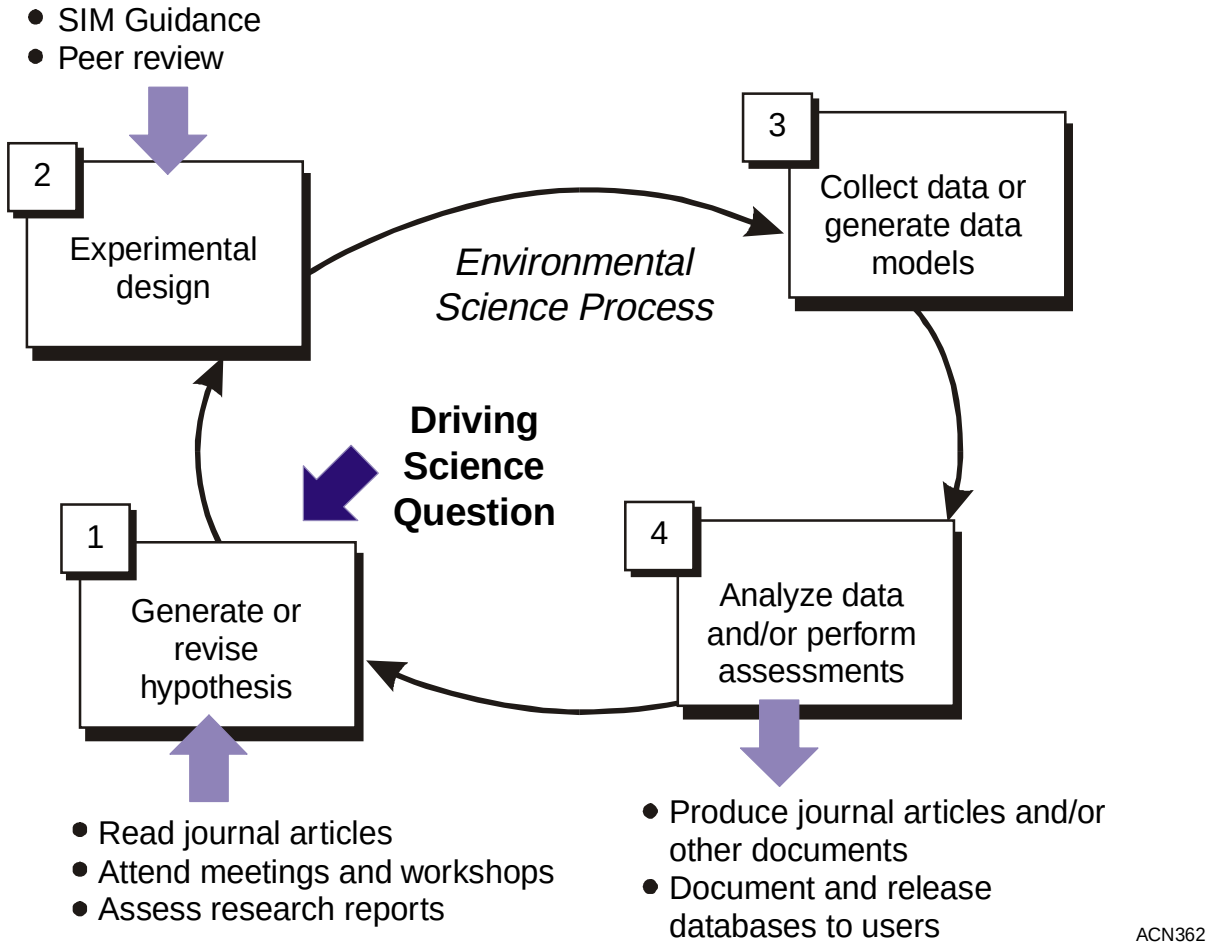


Figure 4. ORD Environmental Science Process

The functional requirements for ORD's scientific systems are illustrated below in Figure 5. The figure uses the case scenarios, discussed in the Preface, to illustrate the types of questions that typical system users might ask during assessment activities, and outlines the three major functions of a scientific information management system. Although the scope of the example is limited, it serves to define the types of information that a distributed ORD information system is required to provide in order to meet its users' requirements.

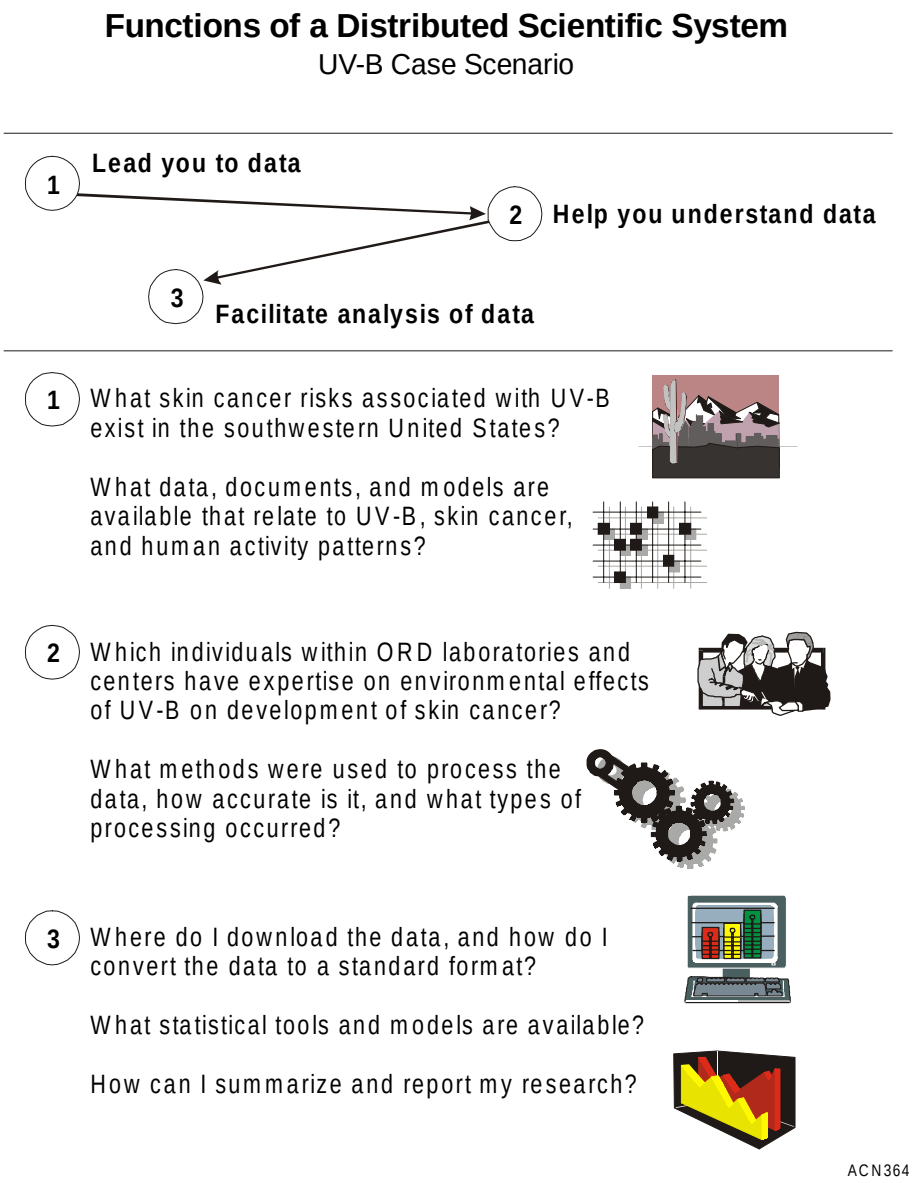


Figure 5. Functional Requirements for ORD's Scientific Systems

The foregoing sketch of the research process and the functional requirements for scientific information systems provides a context for consideration of user requirements for scientific information.

5.2 Groups of Users

5.2.1 ORD Research Scientists

5.2.1.1 Background

A typical research scientist has an extensive background in science but is not likely to be an expert in information technology. Supporting these scientists are data administration staff who store, manage, and make data available to the research scientists. A variety of computing hardware, software, and data management procedures are currently used by scientists and data administration staff.

5.2.1.2 Requirements

Research scientists' primary activities are planning, implementing, analyzing, and reporting. Planning involves designing the research program and conducting associated logistics work. Implementation entails the field and laboratory measurements, sample collection and analysis, and processing of information required to conduct the scientific study. Once collected, data is assessed, used in scientific simulations, analyzed, and used to produce various informational reports. Primary requirements for ORD research scientists include:

Planning/Design: Data previously collected must be available for review prior to conducting new field operations. During the research planning phase, there is a particularly great demand for ad hoc access to existing data and descriptive information (metadata) to facilitate its interpretation. Software to support data exploration, visualization, and analysis must be provided. Population of informational database reference tables (such as state or county), is helpful to research teams in that they expedite the data entry process and ensure data quality.

Data Access: Data collected should be available internally to the research scientists as quickly as possible.

Data Integrity: Data should be handled appropriately and consistently to prevent corruption of data and loss of integrity. The history of data manipulations should be available to the scientist and reproducible. Information accuracy must be ensured to avoid widespread ramifications including misleading research efforts, incorrect estimate of trends, skewing of results, and possible misdirection of U.S. environmental policy.

Modeling, Visualization, and Analysis Tools: A variety of modeling, visualization, and analysis tools should be available for conduction of scientific studies. These tools should be readily accessible and usable by scientists; information management systems should be easily accessible and provide easy data interface consistent with tool requirements.

Documentation: Data must be consistently recorded and documented over an extended period of time. Data administration policies and procedures for documentation and distribution of data and metadata should be established to facilitate defensible analyses over varying temporal periods.

Queries and Reports: Reports must be generated to produce planned research products and represent the state of research for which each group is responsible. These reports are generated from the data and metadata created from gathered information and should be automated as much as possible.

Methods, Indicators, and Reference Databases: Methods, indicators, and reference databases should be available for the conduct of scientific studies. These resources should be readily accessible and usable by scientists; information management systems should be easily accessible and provide easy data interface consistent with resource requirements.

IM Research Tools: Information management research tools should be available for conducting scientific studies. These resources should be readily accessible and usable by scientists; information management systems should be easily accessible and provide easy data interface consistent with resource requirements.

5.2.1.3 Use of ORD IM in Meeting Research Scientists' Requirements

ORD IM supports the Research Scientist in developing operational information management systems in a distributed environment. While data and metadata owned by each group is maintained independently, scientists can leverage existing information located in ORD central databases, as well as data owned by other research groups. It is anticipated that this environment will include a number of analytical tools designed to facilitate access to ORD data and metadata. In addition, ORD IM will institute a series of data documentation and manipulation standards to improve the quality assurance rating of Agency data. In order to meet the requirements listed, ORD IM will provide:

Networked information systems that work on a distributed basis and collectively meet Agency requirements.

Existing systems will be evaluated (Environmental Information Management System [EIMS], Environmental Monitoring and Assessment Program [EMAP], Consolidated Human Activity Database [CHAD], Total Human Exposure Risk Database and Advanced Simulation Environment [THERDBASE], etc.) and additional requirements collected to determine overlap and voids in ORD system functionality. Work completed or in process should be leveraged appropriately to conserve ORD financial and time resources.

Networked access to modeling and analysis systems that work on a distributed basis and collectively meet Agency requirements.
Existing systems (Third Generation Air Quality Modeling Systems [Models-3], THERDBASE) will be evaluated and additional requirements collected to determine overlap and voids in ORD system functionality. Work completed or in process, such as the Multimedia Integrated Modeling System (MIMS), should be leveraged appropriately to conserve ORD financial and time resources.

Uniform metadata standards complying with Federal Geographic Data Committee (FGDC) standards.
EIMS is a prototype for a comprehensive, environmental information metadata database and has a legacy of metadata storage at the dataset level. Metadata standards used in the EIMS system should be evaluated and FGDC-compliant requirements should be integrated; the relationship between SIMCorB and EIMS should be defined. Issues surrounding the conversion of existing metadata to the new standard format must be addressed.

Uniform data elements naming standards consistent with the Environmental Data Registry (EDR) naming structure.
EIMS will most likely be integrated with the EDR, which will replace the current dictionary structure. Implementation of EDR data elements creates a standard naming structure for information within the Agency. Currently, EDR has more than 230 data elements standardized within the database.

Taxonomic and chemical standards and methods.
Taxonomic and chemical standards and methods should be researched and adopted.

Project documentation and data administration standards.
Specifications for project documentation standards and procedures should be discussed. Scientific and systems development staff should collaborate with the OMIS team and program/project managers to formulate appropriate standards prior to coding changes.

Archival procedures and processes with the objective of providing long-term scientific data archives.
Detailed procedures for archiving data must be defined; the type of data to be archived and the longevity of the dataset should be addressed. Metadata collection and documentation requirements should be enhanced in order to facilitate long-term data usability. Metadata requirements should be collected with the goals of:

- Ensuring that analysts understand original user requirements and the appropriate use of the data.
- Verifying the integrity of source data and providing comprehensive quality assurance information.

- Maximizing benefits while minimizing resources.
- Allowing data to be retrievable in a useful way for a period of 20 years following original data collection and documentation.

Uniform search and retrieval capabilities for scientific information and archives.
A common search interface providing keyword, temporal, and geographic search of ORD scientific information should be developed. A common keyword lexicon should be developed, and integration of a textual search engine should be assessed. Geographic search parameters should include Country, State, County, Zip Code, Municipality, and GeoArea. Temporal searches should be developed for series of dates, including Sampling and Publication dates.

Uniform assessment capabilities for scientific information and archives.
A common interface should be developed that provides general assessment capabilities for ORD scientific information. The interface should permit online evaluation of data with online analytical processing (OLAP) capabilities, including sorting, reporting, graphing, calculations, and trend analysis.

Improved tools for a research area.
The environment should provide various tools for a research area. Tools should be readily accessible from a central point of access although they may physically reside at various locations. The environment should provide data interface consistent with tool requirements.

Improved compliance measurement methods, indicators, and reference databases.
The environment should provide access to methods, indicators, and reference databases. Methods, indicators, and databases should be readily accessible from a central point of access although they may physically reside at various locations. The environment should provide data interface consistent with resource requirements.

Improved tools for measuring, modeling, assessing, and restoring ecosystems.
The environment should provide access to measuring, modeling, and assessment resources. Measuring, modeling, and assessment resources should be readily accessible from a central point of access though they may physically reside at various locations. The environment should provide data interface consistent with resource requirements.

Improved exposure/effects assessment, better assessment methods, and restoration options.
The environment should facilitate access to and analysis of data used to perform exposure/effects assessment. Data and decision-making tools should be readily accessible from a central point of access although they may physically reside at various locations.

Improved risk assessment, ecosystem management, decision-making, monitoring, and models.

The environment should provide enhanced monitoring and modeling capabilities used in risk assessments, ecosystem management, and decision-making. The environment should facilitate access to and analysis of resources used to perform risk assessments, ecosystem management, and decision-making activities. These resources should be readily accessible from a central point of access although they may physically reside at various locations.

5.2.2 ORD Executives and Managers

5.2.2.1 Background

Executives have extensive regulatory and decision-making experience but might not be experts in either environmental sciences or information technology. Their IM applications will be at the policy level. Typically, they will use personal computers for data analysis or presentation development. Requests for data will often be made by phone, or other personal contact, rather than through direct computer access.

5.2.2.2 Requirements

This user group will primarily be interested in successful implementation of new policies and standards, as well as support from ORD researchers and staff. For information systems, this group will be accessing summary level information. Requirements include:

Researcher Support: Earn the support of research scientists regarding new policies and standards. Facilitate introduction of new policies by redefining responsibilities within the organization.

Cost Efficiency: Provide state-of-the-art information management systems at, or below, budget. Minimize duplication of effort, and encourage information sharing between ORD staff. Leverage existing development and research.

Cooperative Inter-office Efforts: Promote interaction between various agency activities to minimize resources and maximize information sharing. Support Agency initiatives, including the Freedom of Information Act.

Risk Assessment/Risk Management: SIMCorB, in collaboration with the ORD Science Council, should identify a major science program area involving multiple ORD organizations and covering the complete risk assessment/risk management process, and lead the development of a demonstration project to meet program IM needs while testing key concepts and technology.

Accountability: Provide accountability for all ORD projects through all phases of the project life cycle.

Publicity: Initiate outreach programs to further Agency and public awareness of ORD activities.

5.2.2.3 Use of ORD IM in Meeting These Requirements

ORD IM will foster a series of networked information systems that collectively meet user requirements for data access, thereby enhancing data quality and minimizing duplication of work. Agency support, budget, and staffing issues are addressed below.

Metadata cataloging responsibilities of all ORD personnel should be clarified.
New metadata collection and documentation requirements necessitate redefining several roles within ORD to meet these standards. The relationship of SIMCorB to various ORD positions, including ADP Coordinators, QA Managers, and Branch Chiefs, should be explored. Create a new position, Scientific Information Coordinator, to specialize in metadata storage and documentation.

All projects within ORD that relate to scientific information management, both present and future, should be assessed and integrated where appropriate.
In order to adopt a long-term strategy with the goal of leveraging individual capabilities from multiple systems, efforts among all ORD offices should be coordinated. The functionality and current status of all ORD projects should be researched and documented.

Architectural and systems engineering issues should be defined.
Development of a cohesive, ORD-wide environment utilizing existing development presents unique architectural issues. System design, implementation, maintenance, and future upgrade concerns should be addressed.

Candidate systems for Reinventing Environmental Information (REI) constituency should be identified.
SIMCorB should solicit sponsorship from REI for an ORD system. (REI's focus is on incorporating data standards and electronic reporting into EPA's national information systems.) Existing systems should be evaluated and the best example of a national ORD environment marketed to REI as a national information system. EIMS may logically fit into this category.

The role of the SIMCorB Executive Advisory Committee in relation to working-level SIMCorB members should be defined.
SEAC was recently formed to allow executive-level staff to participate in the SIMCorB effort while accommodating their unique schedules. The role and responsibility of SEAC should be further clarified and defined.

Improve regulatory/management actions and provide better environmental protection.
Scientific information management should aid ORD Executives and Managers in regulatory decision making, thereby improving overall environmental protection.

Ensure that laboratory and center goals, objectives, and sub-objectives are met in accordance with the GPRA Planning Hierarchy.
The environment should aid ORD laboratories and centers in meeting various objectives through improved scientific information management. The environment should provide functionality in accordance with IT Outcomes outlined in the GPRA outcome pyramid.

5.2.3 Policy Makers-Decision Officials in Government and Industry

5.2.3.1 Background

This category of users, including their staff or technical advisors, employ scientific and technical information to reach decision on regulatory strategy and compliance processes, site specific problem assessments and selection among alternatives for remediation or pollution prevention actions, and assessment of the effectiveness of environmental management programs. A major focus is on the identification of accurate, useable information and on synthesis and integration of diverse information sources to support decisions.

5.2.3.2 Requirements

Information management requirements to support performance of these roles include:

- information on the location and accessability of scientific and technical information related to specific stressors and related environmental processes.
- methodologies and technical capacity to link diverse information resources
- information on data quality and integrity
- information on the spatial and temporal context of data and technical information
- support for interpretation of data and associated contextual information.
- methods and technical capacity for “scaling” information for different levels of application.

5.2.3.3 Use of ORD IM in Meeting These Requirements

The ORD scientific information management strategy includes key resources to help meet these requirements. These include:

- broadly focused communication of the existence of a fully accessible directory-catalog-dictionary system (SIMS) to support identification and access to key scientific and technical information.
- adoption of standard spatial and temporal identifiers for all data and information.
- linkage of SIMS to bibliographic resources for scientific and technical reports and related documents.
- linkage of SIMS to information on analytical methods, assessment protocols, synthesis methodologies, etc.
- linkage of SIMS to models, their metadata, and their documented model outputs.

5.2.4 Stakeholders in Community-based Environmental Problem Solving

5.2.4.1 Background

The stakeholders involved in community-based environmental protection efforts includes a broad array of public officials, concerned citizens, business and industrial leaders, and technical advisors (to cite but a few categories of participants). The specific mix of these stakeholders and their level of sophistication in utilization of scientific and technical information in problem-solving vary with each situation.

5.2.4.2 Requirements

In spite of this operational diversity, some common themes in requirements appear to be present. These include:

- the need for information to be presented in multiple formats for use by different audiences.
- the need for information on the location, accessability and usability of relevant scientific and technical information.
- the need for information on technical assessments and problem solutions in other locations.
- the need for usable information (i.e., interpreted data) communicated in user friendly language.
- the need for information on assessment methodologies and protocols.
- the need for technical capability for synthesis and integration of data and information from diverse sources.

5.2.4.3 Use of ORD IM in meeting these requirements

The ORD scientific information strategy will help support these requirements in the following manner:

- through the SIMS directory-catalog-dictionary system provide full access to a "library" inventory of available scientific and technical information (including reports; other documents) relevant to specific classes of environmental problems.
- identification of relevant data sets/data bases and associated metadata to support site-specific integrated assessments.
- direct access to guidance and methodologies for risk assessment and priority setting.
- information on processing tools (e.g. models and data analysis software).

5.2.5 External Scientists

5.2.5.1 Background

The scientific community external to ORD comprises yet another extremely important but diverse constituency for ORD scientific information. This constituency includes academic researchers and analysts, scientists in other Federal, State and local government and international agencies, EPA regional and program office scientists and technical managers, as well as scientists in industry.

5.2.5.2 Requirements

Although this class of users reflects significant diversity, some common requirements appear to be present. These are:

- information on current and proposed scientific research conducted or sponsored by ORD.
- information on, and access to, reports, other documents, and data developed through research previously sponsored or conducted by ORD.
- information on ORD/EPA scientific information management processes, including external access/linkages.
- assistance in linking diverse information resources to support synthesis and integrated analyses.
- connection among workgroup and research team members sharing methods, data and interim products while working on common problems.

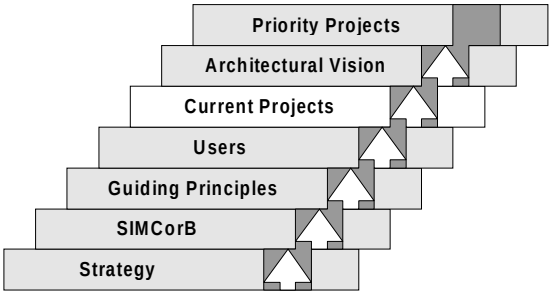
5.2.5.3 Use of ORD IM in meeting these requirements

The ORD scientific information strategy addresses these following requirements:

- full development of OMIS and linkages with SIMS will provide access to information on current and planned research conducted or sponsored by ORD.

- the bibliographic capability being designed for SIMS will permit identification of research reports, documents and data produced through past ORD sponsored research, and development efforts.
- SIMCorB will stimulate and coordinate a systematic effort to enhance two-way collaboration with external scientific groups, including identifying opportunities for integrating scientific information management capabilities of EPA with external systems.

6 ORD Laboratory and Center IM Responsibilities and Projects



Following the explanation of different users and their needs in Section 5, this section presents a list of the projects in which these users are engaged. While not comprehensive, the list does convey the remarkable diversity of projects underway at ORD. The analysis of current SIM related activity in regard to the dimensions of system architecture discussed in Section 7 demonstrates potential opportunity for leveraging current investments, as well as highlights voids in current efforts. This section also relates IM integration in ORD laboratories and centers; IM responsibilities of ORD units; scientific information management policies, standards, and cross-cutting procedures of SIMCorB.

6.1 Importance of IM Integration Across ORD Laboratories and Centers

As discussed above in relation to the ORD Strategic Plan, environmental risk assessment/risk management will call upon all of the scientific resources of ORD. As depicted in the case scenarios, individual assessments will involve the scientists and analytical capabilities of several ORD sites simultaneously, many of which are frequently linked with external collaborators. The sequencing of research efforts—from understanding dose-response, to exposure assessment or risk reduction strategies— will frequently entail multiple uses of data or analyses. Failure to establish the capability for integration of data, information, and research results across ORD laboratories and centers will seriously impede accomplishment of ORD's mission responsibilities.

6.2 Information Management Responsibilities of ORD Units

Effective management of scientific information must consider the responsibility for effective management of ORD resources assigned to each organization. The management strategy, reflected in this plan, is to identify the IM components of existing management systems and assure that they are properly defined and linked to the management control processes—from planning through resource allocation, implementation, documentation, publication, and archival. ORD laboratories and centers have responsibilities for their respective organizations; however, SIMCorB is responsible for coordination among the different organizations. Furthermore, the establishment of a laboratory-level science IM coordinator/liaison position would allow coordination of laboratory-level IM activities and contribute to continuing IM

evaluation and improvement. This concept will be fully assessed early in 1999 for consideration by the Science and Executive Councils.

6.3 Development and Implementation of Scientific Information Management Policies, Standards and Cross-Cutting Procedures Adopted by SIMCorB

The laboratories and centers within ORD have responsibility for ensuring effective development and implementation of the scientific IM policies, procedures, and standards adopted by ORD as a whole. In support of this responsibility, SIMCorB proposes assessment by each operating organization of its capacity for effective involvement in SIM policy development and implementation. In addition, an ORD-wide training and education project on IM concepts and issues is included in the priority action projects for 1999.

6.4 Unit or Program-Specific Development Projects Linked to Scientific Information Management

Numerous projects with implications for scientific information management are either planned or now underway in each of the ORD organizational units. Most of these projects touch on cross-cutting information technology and/or scientific information management issues. Taken together, these projects provide a baseline for evaluating current opportunities for improvement and planning the direction of future scientific information management initiatives.

6.4.1 Office of the Assistant Administrator—ORD

6.4.1.1 Environmental Monitoring and Methods Committee (EMMC)

An EPA-wide committee representing ORD and program offices that seeks compatibility and standardization where possible of monitoring techniques, analytical methods, and models and data classifications employed in EPA regulatory and monitoring programs.

6.4.1.2 Models 2000 Steering/Implementation Plan

An EPA-wide committee to draft guidelines on model acceptability criteria and peer review for modeling applications.

6.4.1.3 Supercomputing Executive Committee (SEC)

An EPA-wide committee focused on planning for future supercomputing, model data management, and visual needs.

6.4.2 Office of Resources Management and Administration (ORMA)

6.4.2.1 Year 2000 Project

An initiative to identify and correct date-linked software vulnerabilities.

6.4.2.2 OMIS Expansion-Science IM Link

A proposed project to realize the ORD strategic goal of tracking all ORD sponsored research projects in OMIS; including documentation and standards on scientific IM issues.

6.4.2.3 Chief Information Officer (CIO)

Position established with responsibility for ORD-wide management of information resources, development of sound, integrated information technology architecture and information management processes, and issuance of ORD-wide policies and procedures governing information technology and information resources.

6.4.2.4 Management Information Systems Staff (MISS)

Centralized information technology support staff with oversight responsibilities for OMIS.

6.4.3 National Center for Environmental Assessment (NCEA)

6.4.3.1 Environmental Information Management System (EIMS)

A metadata management and data analysis system to support integrated risk assessment.

6.4.3.2 National Health and Nutrition Examination Survey (NHANES)

This is a nationwide project led by the Centers for Disease Control (CDC) in collaboration with EPA and other Federal agencies to develop a statistically representative snapshot of environmental health conditions of the general U.S. population and significant sub-population; NCEA's focus is on risk assessment.

6.4.3.3 EIMS Data Types and Source Project

A project to develop the EIMS application to incorporate references to all information types relevant to NCEA personnel including pointers to sources.

6.4.3.4 Bibliographic Software Assessment

A project to define requirements and specifications for bibliographic software needed by NCEA personnel, and the implementation of a central repository for bibliographic entries and links to document storage files.

6.4.3.5 Collaborative Work Process Infrastructure

A project to create an internal communication tool for scientific collaboration; including capability to work collaboratively on documents, manage documents, and track and share correspondence.

6.4.3.6 Risk Assessment Models and Software

A project to develop a suite of software and models needed to complete risk assessments; including capability to distribute via CD-ROM or the Internet.

6.4.3.7 Integrated Risk Information System (IRIS) Modernization

A project to establish IRIS in an Oracle systems environment and establish access via the Internet.

6.4.3.8 Geographic Information System (GIS)/Visualization Software

A project to expand the suite of GIS applications available to NCEA personnel to support risk assessments.

6.4.3.9 Information Tools Technology Training

A project to make training available for all NCEA personnel on available IT tools; including Internet use, word processing and spreadsheet applications, database structures and use, GIS and visualization software, bibliographic software, etc.

6.4.3.10 Scientific Systems Requirements Analysis

A project to conduct a complete requirements analysis of the scientific software and functionality needed to support NCEA's scientific modeling and analysis needs.

6.4.3.11 Enhanced Public Access Task Force—Environmental Monitoring for Public Access and Community Trading (EMPACT)

A national EPA project to promote real-time public availability of environmental monitoring information in user-friendly formats; NCEA-Cincinnati task-level leadership.

6.4.3.12 EPA Risk Assessment Forum

An EPA-wide standing committee of senior scientists which develops human health and ecosystem risk assessment guidelines.

6.4.3.13 Global Change Research

An ongoing, interagency program of global climate change research, wherein NCEA provides leadership in integrated assessment activities.

6.4.3.14 Information Resources Development Staff (IRDS)

A central NCEA staff office with responsibilities for organization-wide leadership, coordination and management of information technology and information management processes.

6.4.4 National Center for Environmental Research and Quality Assurance (NCERQA)

6.4.4.1 Scientific IM Links to ORD Competitive Research Grants—Science to Achieve Results (STAR)

A proposed project to tie scientific IM documentation, standards, and accessability to projects funded under the STAR program and tracked within OMIS.

6.4.4.2 EIMS Data Quality

Continuing technical and policy support on QA issues related to EIMS development.

6.4.4.3 Enhanced Public Access Task Force—EMPACT

Project described at 6.4.3.11; NCERQA task-level leadership.

6.4.5 National Exposure Research Laboratory (NERL)

6.4.5.1 Electronic Notebook Pilot Project

A project to assess the feasibility of, and technical requirements to support, the use of electronic notebooks for documentation and archiving of scientific notes and instrument readings.

6.4.5.2 Human Exposure Information Management

A project to develop and validate systems to make NERL exposure data, methods, and models readily available and to integrate human exposure platforms, data, and models into the NERL EIMS.

6.4.5.3 Total Human Exposure Risk Database and Advanced Simulation Environment (THERDBASE)

A project to upgrade the technology supporting this system for providing databases and models for risk assessment to include Internet communication and expanded links to large exposure databases held by other organizations.

6.4.5.4 Consolidated Human Activity Database (CHAD)

A project to provide an Internet-accessible, user friendly, comprehensive database of all daily human activity surveys, with links to evolving high performance modeling software.

6.4.5.5 Survey of ORD Data Management Systems

A project to extend ORD-wide the NERL pilot survey of information analysis and management software and technology currently deployed in ORD organizations.

6.4.5.6 Capture of Data from Analytical Instruments

A project to assess feasibility and define requirements for capture and transmission of data produced by scientific analytical instruments.

6.4.5.7 Information Management for NERL Endocrine Disruptor Research

A project to create a metadata management capability for NERL endocrine disruptor data and documents.

6.4.5.8 North Atlantic Research Strategy for Tropospheric Ozone (NARSTO)

NARSTO is a public-private international consortium for research on tropospheric ozone; one NERL project will be to assess the NARSTO quality systems management plan and scientific information documentation and archival standards for possible incorporation into, or linking to, ORD systems.

6.4.5.9 UV-B Data Project

A project to recover and install in a relational database UV-B monitoring data collected in earlier pilot activities.

6.4.5.10 Multimedia Integrated Modeling System (MIMS)

Design and development of an open architecture system for creating and applying multi-media models to support ecosystem assessment activities including automated data, metadata, and model management.

6.4.5.11 NERL EIMS

A continuing project to establish a NERL-focused EIMS for metadata management.

6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support

A project to establish EIMS metadata management and analytical support for major regional environmental vulnerability assessments.

6.4.5.13 Third Generation Air Quality Modeling System (Models 3)

A project to provide wide-spread access and use of air quality modeling tools and associated distributed data for state Implementation Planning and ecosystem exposure assessment.

6.4.5.14 National Human Exposure Assessment Survey (NHEXAS)

A multi-region study to better understand the complete picture of human exposure to all types of toxic chemicals through all routes of exposure; major databases planned for release in 2000.

6.4.6 National Risk Management Research Laboratory (NRMRL)

6.4.6.1 Enhanced Public Access Task Force—EMPACT

Project described at 6.4.3.11; NRMRL task-level leadership.

6.4.6.2 Management and Distribution of ORD Scientific and Technical Publications

A project to develop a master ORD publications database by converting and expanding the ORD bibliography of 23,000 publication records for public, searchable access via the Internet.

6.4.6.3 Alternative Treatment Technology Information Center (ATTIC)

A continuing program of information dissemination on remediation technology deployed at Superfund sites.

6.4.6.4 Ground Water Models

A continuing program capability of technical advice and application of ground water models for environmental assessment.

6.4.7 National Health and Environmental Effects Research Laboratory (NHEERL)

6.4.7.1 National Health and Nutrition Examination Survey (NHANES)

See Section 6.4.3.2 for project description—NHEERL focus is on analysis of health effects.

6.4.7.2 Environmental Monitoring and Assessment Program (EMAP)

Science IM projects to create and maintain a metadata management and integrated data analysis capability to support broad scale, integrated, environmental assessments and long-term monitoring of status and trends of environmental resources.

6.4.7.3 Quantitative Structural Activity Relationship (QSAR) Modeling Systems

A series of projects at the NHEERL-RTP facility to perform quantitative chemical structure - biological activity relationship analysis for multiple health endpoints.

6.4.7.4 ECOTOX

An ECOTOXicological database and an associated QSAR-based modeling system, known as the Assessment Tool for Ecological Risk (ASTER). ECOTOX represents an integration of the Aquatic Toxicity Information Retrieval (AQUIRE), the Terrestrial Plant Toxicity Base (PHYTOTOX), and the Terrestrial Wildlife Toxicity Database (TERRETOX), three existing NHEERL databases that contain ecotoxicity information for aquatic life, terrestrial plants, and wildlife, respectively.

6.4.7.5 EPA/International Agency for Research on Cancer (IARC) ARC Graphic Activity Profile (GAP)

A graphic activity profile database on genetic and related effects of potential human carcinogens developed by NHEERL in collaboration with IARC.

6.4.7.6 Information Management for NHEERL Endocrine Disruptor Research

A project to create a metadata management capability for NHEERL endocrine disruptor data and documents, as well as a graphic activity profile database for endocrine disruptor analogous to EPA/IARC GAP.

6.4.8 Linkage of Current Scientific IM Activities to Planned SIM Architecture

Table 1 presents a distribution of current scientific IM-related projects and information management responsibilities organized by the major categories of planned SIM architecture (see Section 7). This summary chart serves as a technique for transition in perspective from the current more or less fragmented array of current efforts to a more focused, targeted management strategy.

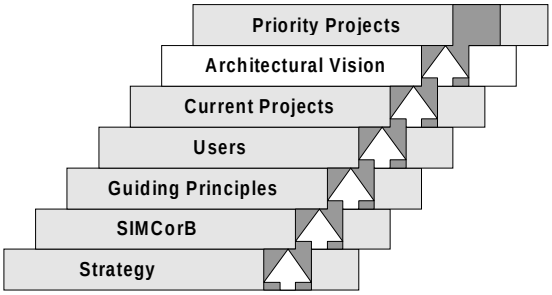
Many of the current projects include elements that could be classified in several categories of the planned architecture. For the purpose of this analysis, classification of each project into a single category is based upon the dominant character of the project.

Review of the distribution of current activities highlights the difference in current engagement in SIM activities across the major ORD organizations. Of particular note is the limited activity in the organizational realm. Alternatively, significant clusters of current activity in the network/systems development area, as well as processing functions, highlight the potential for leveraging investments through improved coordination.

Table 1. Distribution of Current IM Projects by Architecture Category
*Listing includes current projects and information management staffing.

Architecture Dimension	Office of ORDAA	ORMA	NCEA	NCERQA	NERL	NRMRC	NHEERL
Policies, Procedures & Standards	6.4.1.1 6.4.1.2		6.4.3.10 6.4.3.12	6.4.4.1 6.4.4.2	6.4.5.5 6.4.5.8		
Data and Information (including data collection)			6.4.3.2 6.4.3.3		6.4.5.2 6.4.5.4 6.4.5.9 6.4.5.14	6.4.6.2	6.4.7.1
Processing Functions			6.4.6.4 6.4.3.6 6.4.3.7 6.4.3.8 6.4.3.13				6.4.7.3 6.4.7.4 6.4.7.5
Network (including external information transfer)		6.4.2.1 6.4.2.2	6.4.3.1 6.4.3.5 6.4.3.11	6.4.4.3	6.4.5.1 6.4.5.3 6.4.5.6 6.4.5.7 6.4.5.10 6.4.5.11 6.4.5.12 6.4.5.13	6.4.6.1 6.4.6.3	6.4.7.2 6.4.7.6
Organization	6.4.1.3	6.4.2.3 6.4.2.4	6.4.3.14 6.4.3.9				

7 ORD Scientific IM Architecture Vision



The previous Section introduced the ORD laboratory and office projects and responsibilities. This section describes the components of the information management architecture needed to fully support integrated scientific information management. As illustrated in the hypothetical case scenario, scientific IM will require a comprehensive architecture to provide the design linkages among major ORD systems and capabilities, to articulate requirements for internal and external communication, and to provide a framework for coordination of SIM projects across ORD.

7.1 Introduction

The ORD Scientific IM Architecture provides the basis for developing a distributed information management system that leverages investments in existing technology, accommodates rapid and continuous changes in technology, and promotes information sharing between the Office, Agency, and general public. Four separate, but interrelated, architectures make up the IM environment:

- Data and Information Architecture
- Processing Function Architecture
- Network and Desktop Architecture
- Organizational Architecture

The development of these architectures is guided by system design criteria and operating requirements emanating from ORD information management policies and administrative management processes. An internally consistent body of IM policies, procedures, and standards constitutes almost an additional separate architectural dimension of the SIMS. Many of the required policies and procedures are already in place; however, priority attention must be devoted to definition and adoption of revised policies in ORD research project approval and documentation, records management, archiving, and other areas to establish criteria for system design.

7.2 Architectural Overview

The system architecture depicted below in Figure 6, *Architecture Overview*, illustrates the scope of ORD scientific IM requirements. The architecture overview diagram shows the interrelationship among three major architectural components: 1) the information that ORD manages; 2) the systems that process the information and perform functions; and 3) the supporting internal and external communications linkages. Underlying this technology architecture is an organizational architecture that contains the policies, procedures, responsibility structures and staffing needed to support the management and sharing of scientific information across ORD and with various external users as envisioned in the ORD mission.

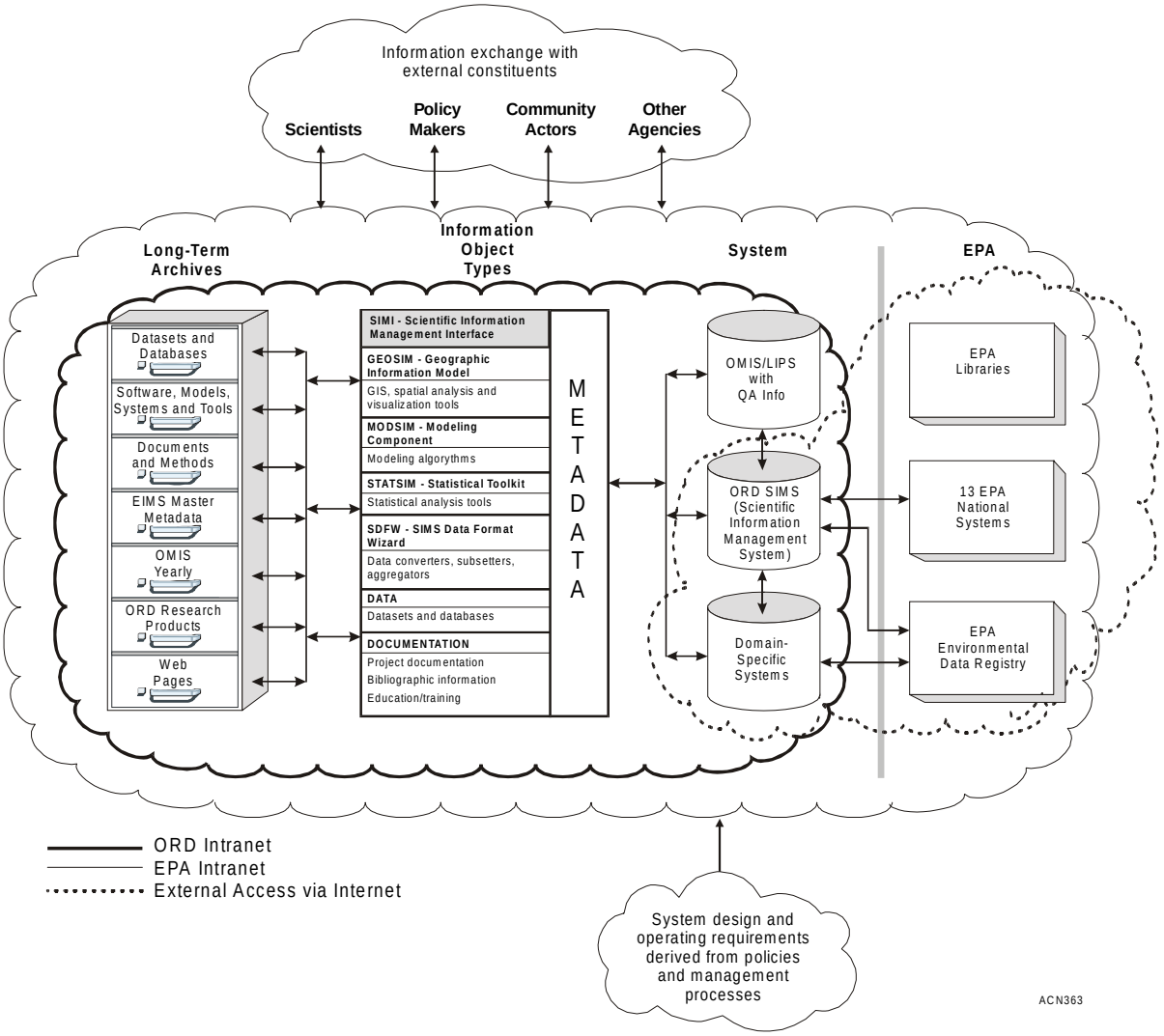


Figure 6. Architecture Overview

7.2.1 Virtual Repository of Current and Archived Data and Information

ORD manages various types of information, or Information Object Types, such as data, metadata, tools and documentation. The distributed architecture provides comprehensive access through selected ORD systems to all available ORD information and resources housed in one or more distributed servers. This virtual repository of object types promotes sharing, reuse, and analysis of ORD's substantial information resources and minimizes duplication of effort within the Office. Examples of these object types are described in Table 2. *General Categories of Information Objects*.

Object types deemed appropriate for inclusion in ORD's long-term archives are archived according to specified procedures at periodic intervals. Information stored in ORD's long-term archives remains available to system users, with the potential for updating through new versions of a data set or other object, as indicated by the double arrows between Long-term Archives and Object Types in the diagram.

Table 2. General Categories of Information Objects

Information Object - Data	Examples
Dataset repositories (collections of data not residing in a formal database management system)	Spatial Datasets (GIS) Collections of Air Quality Modeling Study Results Southern Oxidant Study Data Archive Web Pages (HTML datasets)
Databases (collections of data residing in a database management system)	EMAP EIMS Analytical Reference Databases Taxonomic Methods Chemical Analytes Agency Databases (TRI, STORET)

Information Object - Metadata	Examples
Scientific metadata for 20-year rule (common components of metadata include summary descriptions, methods, quality assurance information, contacts, temporal and spatial information, and keywords describing contents and geographic extent)	EIMS metadata EMAP metadata

Table 2. General Categories of Information Objects Cont'd

Information Object - Tools	Examples
Education and training materials	Web-based Training (examples: Computational Atmospheric Science, Air Quality Meteorology)
Analytical procedures and methods	BioMediation in the Field System (BFSS) Surface Water Analytic Test Methods Good Automated Laboratory Practices Analytic Test Methods for Drinking Water
Project specific analysis and visualization	Pesticide Assessment Tool for Rating Investigations for Transport (PATRIOT) Package for Analysis and Visualization of Environmental Data (PAVE) VIS-5d–3D analysis and animation Collections of Statistical Analysis Software (SAS) and Graphics Routines used to support project
Scientific software libraries and models	IRIS ATTIC Regional Acid Deposition Model (RADM) Hydrologic Simulation Program Fortran (HSPF) Mesoscale Meteorological Model (MM5) Display software Pesticide Root and Zone Model (PRZM)
Data Converters, subsetters, and aggregators	SAS Procedures Models-3 I/O Applications Programming Interface Library
Modeling and assessment systems	ECOTOX Threshold Software Center for Exposure Assessment and Modeling Software Third Generation Air Quality Modeling System (Model-3) Exposure Models Library and Integrated Model Evaluation System –(new THERDBASE) Better Assessment Science Integrating Point and Nonpoint Source (BASINS)

Information Object - Documentation	Examples
Research program/task documentation	Project Reports Laboratory Notebooks Contract/Cooperative Agreement Reports ORD Research Product Inventory OMIS
QA Information	QA Annual Report and Workplans QA Assurance Project Plans Quality Management Plans QA Assessments
Education and training materials	Training documentation, manuals, etc.

7.2.2 Processing Functions

ORD's major computer systems as shown in the architectural overview diagram have three separate entities:

- OMIS, including the Laboratory Implementation Planning System (LIPS) and quality assurance/ quality control (QA/QC) information.
- the Scientific Information Management System (SIMS).
- domain-specific systems supporting specific science programs such as EMAP, THERDBASE, the Network Analytical Data System (NADS), CHAD, Models-3, and MIMS.

OMIS provides the ability to track all ORD- sponsored research projects and allows users to research financial, project and IM planning, and staffing information for individual projects. SIMS, ORD's integrated scientific information management system, is represented by a future iteration of EIMS. SIMS is envisioned as a means to centralize access to, and facilitate exchange of ORD's scientific and analytical data and metadata within the Agency, as well as with those external to the EPA. ORD's domain-specific systems access a variety of data and generate products that complement the information and project management capabilities of SIMS and OMIS. A double arrow indicates the two-way transmission of information between ORD Systems and Object Types. The critical overarching system design requirement is to assure compatibility among these systems to permit interoperability and efficient sharing of information.

Collectively, the system architecture permits complete access to available ORD information resources by system users. Via SIMI, ORD metadata, data, projects, models, documents, and spatial information may be located, queried, and accessed. Project management data provided by OMIS may be easily referenced at the directory entry level. GEOSIM, the Geographic Information module of SIMI, provides access to GIS and visualization tools. Modeling algorithms are stored in the MODSIM component. Statistical tools may be accessed using STATSIM. SDFW, the SIMS Data Format Wizard, finds, delivers, and converts selected data. Configuration management issues are minimized and an inventory of Agency assets is simultaneously maintained.

7.2.3 Network Communication

ORD systems are supported by a three-tiered communication network which is represented by solid and broken lines on the architecture overview diagram. Systems that require links to information or users outside of the Agency communicate via the Internet, while systems that access information or users across EPA will use EPA's wide area network (WAN). ORD-specific systems, such as OMIS, will operate on ORD's Intranet. Domain-specific systems may

use a combination of the different communication network tiers. All systems may limit access to defined user groups on a case-by-case basis.

7.2.4 Policy and External Users

The overview diagram depicts the policy context of the system, as well as the range of external users of and contributors to ORD scientific information, both of which impose operating requirements on the system.

7.3 Architectural Components

7.3.1 Data and Information Architecture

The data architecture will be used to ensure that data and associated metadata from the various ORD sources will be stored, managed, and retrieved in a consistent manner with known quality. The data architecture will also be used to facilitate information exchange within ORD, EPA, and other programs and scientists. The major types of data and information that ORD manages are shown with examples in Table 3. Critical design issues in the data architecture include definition of core abstract data types or structures (e.g., continuous monitoring data) and design of common routines for data handing regardless of the specific program application, thus setting the stage for interoperability and efficient sharing of information.

Table 3. Categories of Data/Information

Data/Information Type	Data/Information Example
Health-General	Analytical Methods Data Handling and Statistics Physical Data Chemical Data UV-B Bibliographic Data
Hazard Identification	Chemical Monitoring Data Fate and Transport Information Health Effects and Epidemiologic Data Animal Experimental Data Receptor Populations
Dose-Response	General Dose-Response Information Toxicity Data Toxicokinetics Pharmacokinetic Models

Table 3. Categories of Data/Information Cont'd

Data/Information Type	Data/Information Example
Exposure Assessment	Contact Rate Human Activities Building Parameters Product Data Body Weights Lifetime Dermal Absorption Exposure Models Biological Monitoring and Case Studies
Health-Risk Characterization	Regulations and Advisories Risk Models
Ecological-General	Analytical Methods Information Data Handling and Statistics Meteorological Chemical Information Bibliographic Information Habitat Information Biological Indicators Taxonomic Information
Problem Formulation	Exposure Characteristics Information Receptor Information Stressor Source Characteristics Information Ecological Effects Information Assessment Endpoints Conceptual Model Risk Hypothesis Analysis Plan Information on Other Stressors
Characterization of Exposure	Exposure Models Information Monitoring Data Physical Stressor Information Biological Stressor Information Source Characterization Information Data Evaluation and Quality Control Information Bioavailability and Bioaccumulation Information
Characterization of Effects	Chemical Stressor Information Physical Stressor Information Biological Stressor Information Stressor-Response Models Multiple Stressor/Cumulative Effects Models Effects Indices

7.3.2 Processing Function Architecture

The function architecture will be used to ensure that all relevant ORD Scientific IM functions are considered during the system life cycle and that the data architecture takes into account the data and data standards that are needed to perform the designated functions. Also, where appropriate, the function architecture can be used to verify that procedures used by different ORD users are consistent. The completion of the function architecture will require an in-depth review of ORD's functions, policies, procedures, processes, and sub-processes.

7.3.2.1 Major Information Processing Functions

Some of the major processing functions that ORD performs are listed below:

- Data collection or generation
- Data administration and security (data input, validation and verification, access/querying—levels of access, archival, data release)
- Project definition and tracking
- Metadata management
- Development and use of measurements and models
- Human health and ecological risk assessments
- Spatial analysis (GIS)
- Statistical Analysis
- Research activities and links to external sources
- Monitoring (design, measurements, etc.)
- Development of methods and indicators
- Report publication
- Data distribution and public access
- Administration
- Grants management
- Marketing and training
- Bibliographic information management
- Scientific software management

7.3.2.2 Interdependencies in ORD Scientific IM Processing Functions

Figure 7 depicts the interrelationship (and interdependence) among ORD organizations in information structure and information processing functions in regard to ecological risk assessment/risk management. Similar interdependencies are present in regard to human health scientific activities. These interdependencies demonstrate the need for compatibility and interoperability among ORD scientific information management processes and systems.

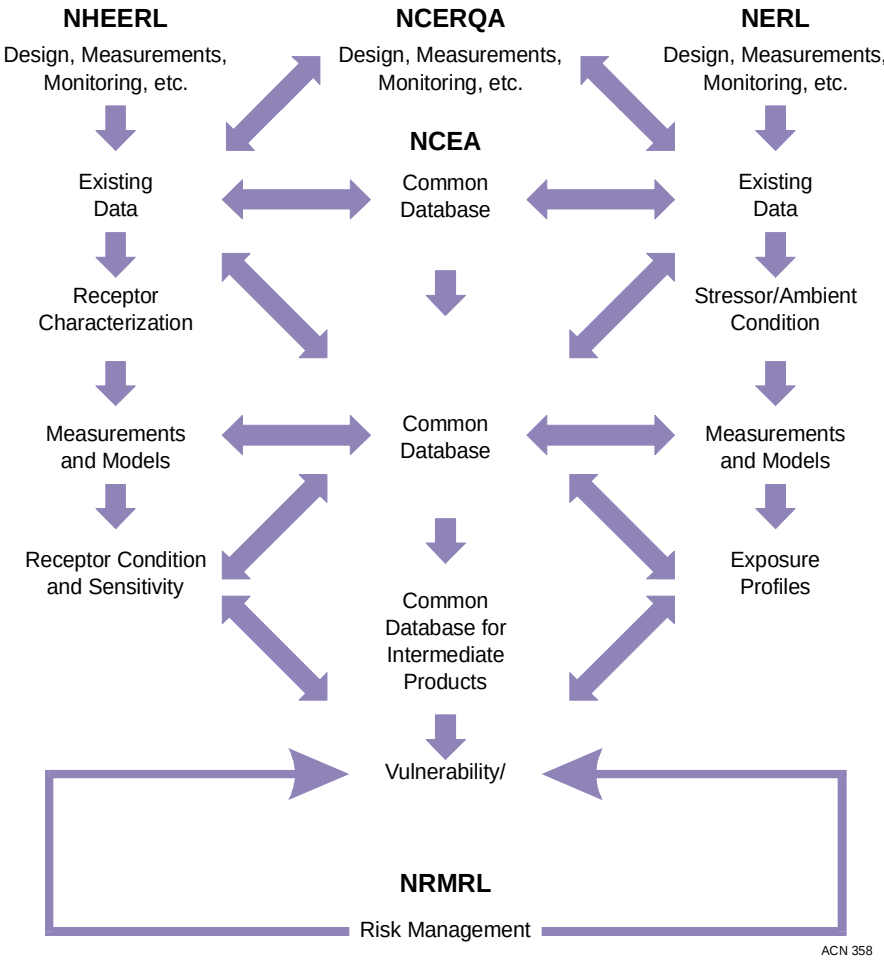


Figure 7. IM Relationships in ORD Ecological Research

7.3.2.3 SIMS Support of Risk Assessment

Figure 8, *Risk Assessment and SIMS*, shows how the risk assessment process outlined earlier uses SIMS' directory, catalog, dictionary, and analytical database (highlighted in Figure 6, *Architectural Overview*) to complete risk characterization. The risk characterization function is divided into several processes, or steps. These include planning, problem formulation, analysis, and publication. At each step, SIMS provides access to information to assist research and assessment scientists, managers, and end-users.

The directory provides summary level identification information about selected directory entries in the SIMS database. During the planning stage, users can quickly assess the available information and determine its relevance to their requirements. Directory information includes general descriptive information about the directory entry, related general and geographic

keywords, and contact information for individuals and organizations associated with the selected directory entry.

The SIMS catalog provides a more detailed level of information about individual directory entries in the SIMS database. Detailed metadata allows users to further evaluate the directory entry and conduct problem formulation activities prior to accessing the data itself. Catalog level information includes collection objectives and methods for the directory entry. Bibliographic references and related directory entries may be recorded, and acronyms are defined.

The dictionary provides information about the data, its availability, and the necessary access procedures. Users may view a snapshot of the actual data, analyze individual data elements in the data dictionary, or access and download the collected data associated with the selected directory entry. Information provided via SIMS' dictionary permits in-depth analysis of the data and eventual publication of risk characterization and risk management results in the SIMS directory.

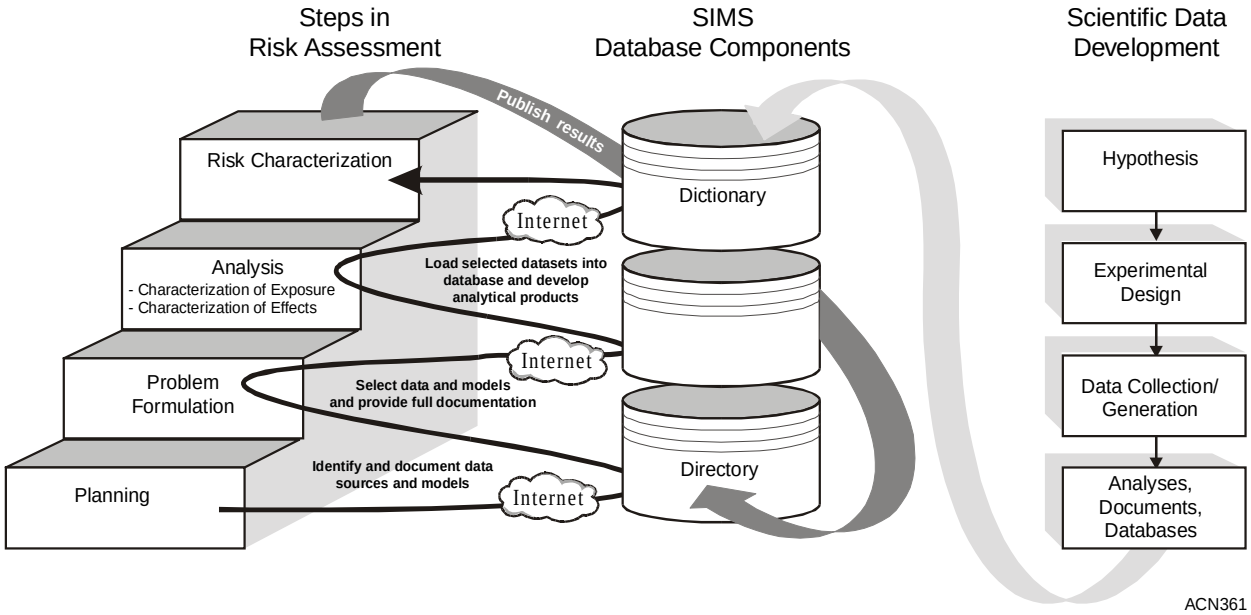


Figure 8. Risk Assessment and SIMS

7.3.3 Network and Desktop Architecture

Figure 9, *ORD Network Architecture*, presents a conceptual view of ORD's technology architecture, including the ORD central computing network, the ORD laboratories and offices computing environment, and public access to ORD information. Figure 9 is representative of the infrastructure that will support scientific IM within ORD, but it is not meant to be inclusive of all new and emerging technologies that might impact scientific IM at ORD.

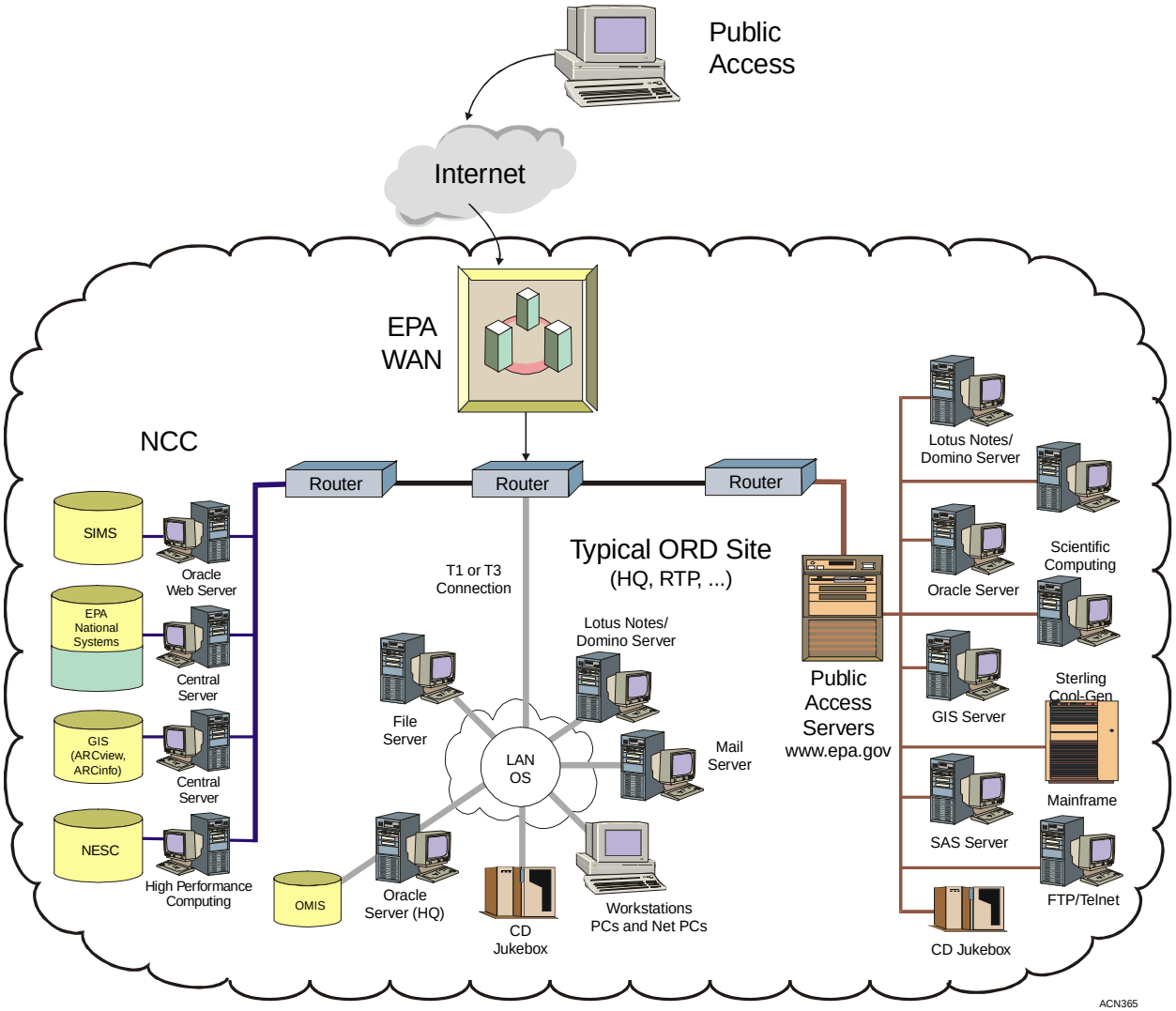


Figure 9. ORD Network Architecture Vision

7.3.3.1 Computing-Communication Structure

Public Access is achieved via personal computers and Internet connections to the public access servers at NCC. Public access is restricted to those applications and information that is explicitly designated for public use; examples of publicly-available systems include EIMS and REI's thirteen national systems.

In addition to Internet access, EPA provides access to various servers supporting EPA standard applications. All applications served by public servers are used throughout ORD in various laboratories and offices; this configuration allows separate laboratories to use applications and information stored centrally in a single repository and permits sharing of both test and production servers. Servers providing access to Lotus Notes/Domino, Oracle applications, Scientific Visualizations, Sterling Cool:Gen, SAS applications, FTP/Telnet, and the mainframe are depicted. A CD Jukebox allowing software CDs to be stored and accessed in a central location is also illustrated.

The ORD site and office computing environment is comprised of divisional local area networks, servers, applications, and workstations. The sample configuration shown in Figure 9 depicts the types of software and hardware implemented at the division level; actual configurations of ORD sites may vary. Servers typically include file and mail servers, a Lotus Notes/Domino server, and an Oracle application server. A CD Jukebox providing access to software CDs at the division level is also depicted. In addition, ORD sites and offices have access to EPA servers and associated applications via T1 or T3 connections.

7.3.3.2 Network and Desktop Architecture Guiding Principles

In order to accommodate rapid and continuous changes in technology, an ORD IM technology architecture will be developed. Key characteristics of this architecture will include the following:

- *Open Systems Preference*. ORD will use industry standard, non-proprietary hardware or software solutions wherever possible. There will be a strong focus on interoperability between various components and integration of legacy systems.

- *Client-Server Approach*. The geographically distributed nature of ORD is ideally suited for a Client-Server approach to systems development. Where data needs to be shared by geographically dispersed users, common databases will be created and maintained on shared servers. For large model datasets distributed data management approaches with common meta databases will be used. Where individuals are working alone or independently, data may be managed on a local server or PC. ORD's user interface will be located on PCs and desktop workstations with Web-enabled database access and local analytical tools for interactive analysis and display of data.

- *Thin Client Implementation.* In client-server applications, a client designed to be especially small so that the bulk of the data processing occurs on the server is called a *thin client*. In contrast, a networked personal computer typically operates as a *fat* or *thick client*, often providing everything except some file storage and printing locally. As the Internet grows in importance, a Web browser will often be the thin client of choice for ORD users.

  Within LAN-type environments, the main benefit of a thin client is ease of maintenance. For fat clients, especially those with poor networking support, installing a new application for each user may entail physically installing the application on each workstation, or modifying client-side configuration options. For thin clients, the maintenance tasks are centralized on the server and need only be done once. Alternatively, thin clients will not be effective for some major functions, such as data intensive visualization applications. The system design must accommodate both options and potential combinations to support specific program requirements.

- *Multi-Layer Architecture with Interchangeable Components.* Because the ORD community has diverse needs and because change is both inevitable and continuous, the ORD Technology Architecture will be separated into multiple layers with well-defined interfaces with the respective components. The goal is to include better, faster, or cheaper components as they become available, with a minimum of cost, disruption, and overall system perturbation.

7.3.3.3 Characterization of the Multi-Layer Architecture

The ORD IM network architecture must be sufficiently robust to support resource intensive modeling and visualization while providing access to widely distributed users and information sources. To support the varied technology requirements, the technology architecture will consist of seven layers. These include the user interface layer, desktop tools and applications, a communications or network layer, a security layer, and the underlying data layers. Figure 10 below illustrates ORD's conceptual IM architecture.

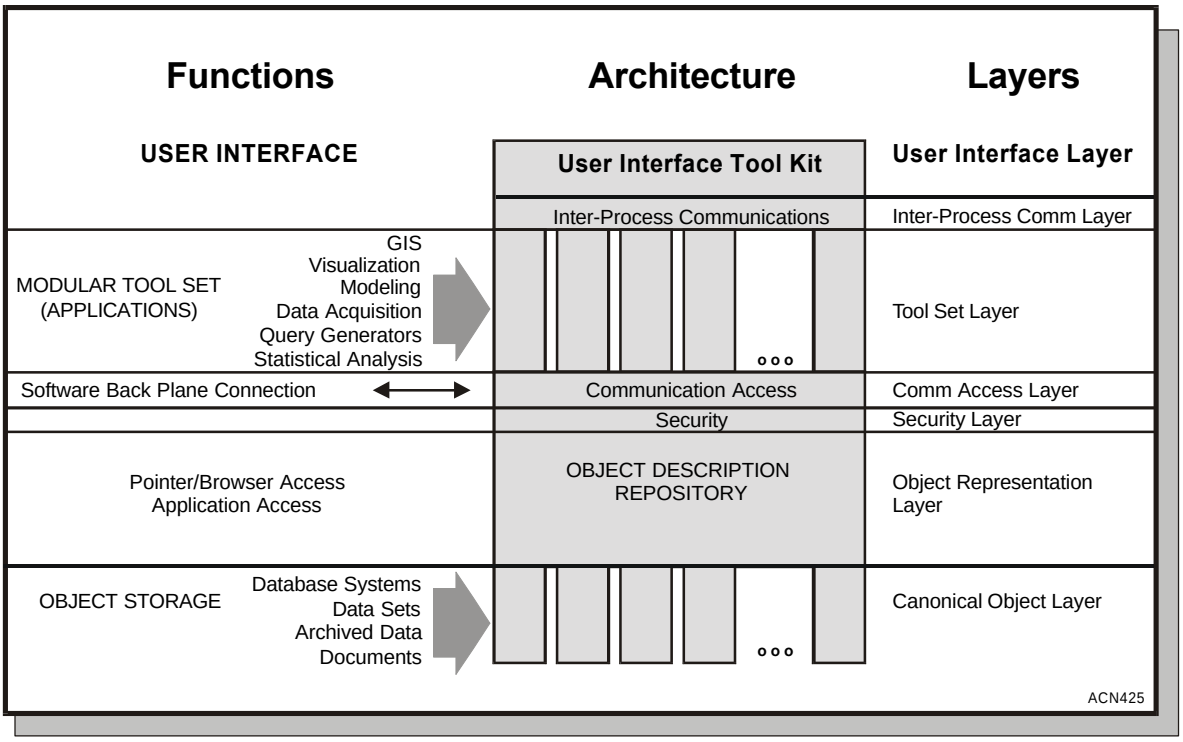


Figure 10. ORD Conceptual IM Architecture

Individual technology architecture components are discussed and examples given in Table 4, *Technology Architectural Components*.

Table 4. Technology Architectural Components

Component Name	Component Description	Component Role	Component Examples
User Interface	"Control panel" that allows users to access a wide variety of tools in a common manner.	Provides mechanisms for developing consistent interfaces for a variety of tools.	Web Browsers JAVA Applets Oracle Forms
Inter-Process Communications Layer	Conduit for information to pass from one tool to another, thereby enhancing capabilities of the system and allowing users to analyze information more completely.	Provides tool-to-tool interchange protocols.	Windows 95, UNIX or NT operating systems; Orb middleware, remote procedure calls.
Tools Layer	Diverse and growing set of information manipulation, data acquisition, and access tools.	Acquires, manipulates, manages, evaluates, and presents information in a wide variety of forms.	Visualization Tools: ~ Graphics/Visualization ~ GIS ~ Analysis Tools Models ~ Statistical Packages Hypermedia Systems Data Loggers Database Reporting Systems Models
Communications and User Access Layer	A tool that provides access to the wide and local area network. Users and systems use these communications and access mechanisms to reach information at other locations. A software backplane connection provides a framework for attaching different packages without disturbing applications already connected.	Provides a path between the different systems Allows many physical implementations of architectural components to be connected to facilitate sharing information and analysis capabilities.	TCP/IP, Internet Dedicated T1, T2, T3 Modems & phone access control cards
Security Layer	Policies, procedures, and technologies to ensure that information is accessible only to proper users. Integrity of information and systems will ensure that information is not altered inadvertently and that any changes to information will be documented.	Protects system, data, and information assets.	Oracle Relational Database Management System (RDBMS) levels of security Oracle Web Server 3.0 Secure Sockets Layer Application Level Security

Table 4. Technology Architectural Components Cont'd

Component Name	Component Description	Component Role	Component Examples
Object Representation Layer	Data Model Manager—a tool that stores models such as entity-relationship diagrams, data flow diagrams, function hierarchies, and database schemas.	Provides information about entities and attributes of a database model and what their relationships are.	Oracle CASE*Designer/2000
	Data Dictionary—contains descriptions, formats, and other basic information for items in a database.	Aids in modeling, developing, and describing the tables and columns stored in a database.	Oracle CASE Dictionary
	Computer-Aided Software Engineering Tools—CASE Tools	Provides access and browsing of the encyclopedia and data dictionary.	Oracle CASE*Designer/2000
	Configuration Manager	Provides mechanism for management of changes, validations, and maintenance of audit trails	
	Directory—a uniform set of descriptions of datasets, data sources, and dataset catalogs with pointers to additional details.	Provides the means to index and track datasets, and provides summaries of these datasets.	Oracle-based directory and catalog system Z39.50 Search Standard
	Catalog—a uniform set of detailed descriptions of a number of datasets and related entities, containing information suitable for making a determination of the nature if each dataset and its potential usefulness for a specific application.	Locates and provides access to detailed documentation on datasets.	Verity Oracle CONTEXT
Canonical Object Layer	A diverse and growing set of data/information objects. This information is stored in various databases and datasets.	Provides organized data for access.	Oracle Databases, SAS Datasets, CD ROMs, Spatial Datasets, Images, Documents.

7.3.4 Organizational Architecture

The organizational architecture required to support a fully functional scientific information management effort within ORD has several key dimensions:

- policy, standards and procedures development and adoption
- SIM roles, responsibilities, authorities and accountability processes
- infrastructure for leadership, oversight, correction and continuous improvement in each ORD organization
- coordinated resource planning for SIM implementation and IT/IM investments
- communication and training

As noted earlier in this Plan, a body of internally consistent policies and administrative procedures, consistently implemented, is essential for an effective scientific information management system.. This requirement is accentuated by the interdependence of ORD scientific activities and need for system compatibility and interoperability discussed above. For example, policies on the release of research data and incentives for "publication of data" as products (e.g. STAR grant program of NCERQA) are needed to allow access by other researchers and use by those involved in risk assessments (e.g. NCEA) and risk management decision making (e.g. NRMRL). Similarly, consistent standards and procedures for data administration (e.g. data input accuracy), system and application development controls (e.g. project management procedures and documentation/reporting pursuant to EPA requirements) and operation and maintenance controls (e.g. software and hardware configuration management) are needed to assure data integrity.

With the establishment of the ORD Chief Information Officer and delegation of authority to the CIO to issue ORD-wide policies and procedures governing information technology and information, the formal structure for adopting and disseminating policy is in place. The establishment of SIMCorB created a mechanism for identification of critical SIM-related policy issues and the development of technically and operationally sound recommendations for consideration by the CIO.Yet to be developed is a structured capability within each ORD operating organization to effectively participate in the SIMCorB process and in the evaluation of SIM policy options and their adoption.

The breadth and complexity of issues associated with scientific information management requires the establishment in each operating organization of a set of defined responsibilities for SIM leadership and coordination and appropriate accountability processes to enhance their performance. An infrastructure for oversight and corrective action are needed (e.g. coordination to retain interoperability) to inform management of the status of SIM implementation and assure continuous improvement. One ORD organization has so organized the SIM function; assessment by other organizations of how these responsibilities should be structured and implemented should be a priority for FY99. This assessment needs to consider

existing roles and capabilities of the ORD IT Coordinators, Quality Assurance Managers, Technical Information Managers, and Records Managers, as well as the added dimensions of SIM discussed in this Plan.

Closely linked to the structure of responsibility for SIM are the issues of adequate staffing and processes for making coordinated investments in IT/IM infrastructure that contribute to the realization of the full capability planned for the ORD SIMS, as well as the specialized mission requirements of each organization for its role in the risk assessment/risk management process.

Organization-wide acceptance of data stewardship values and responsibilities and implementation of SIM concepts must be based upon wide-spread understanding of goals, issues and rationales associated with scientific information management requirements. Establishment of a broad-gauged training and education effort on SIM concepts and the use of SIMS capabilities is an essential near term priority.

7.4 System Development Approach

7.4.1 Modified Life Cycle Concept

In the past, the rigid adherence to a strict Life Cycle Development Approach has not proven to be totally successful in ensuring that operational systems are fully responsive to user requirements. The significant complexity and costs associated with establishing and maintaining information management capabilities have prompted the development of an evolutionary approach that incorporates aspects of a traditional, structured life cycle approach with the benefits of rapid application development (RAD), Computer Aided Software Engineering (CASE) Engineering tools, and prototyping. ORD Scientific IM has adopted an evolutionary approach that encourages end-user participation in all phases of the system design and development process.

Although the traditional phases of the life cycle, listed below, are invoked in the evolutionary process, each phase is limited to a specific functional or technical area. The design process occurs in rapid series with an evolutionary development process that may reuse previously developed system components. ORD users will be able to employ the “proof of concept” systems to perform limited real-world functions and, with increased user feedback, future iterations of the systems will become more robust.

While individual projects are proceeding through the life cycle, SIMCorB will also initiate a global, top-down analysis of ORD’s IM framework to ensure that individual projects will fit into an overall IM infrastructure.

7.4.2 System Life Cycle Phases

SIMCorB's Sub-Group organization for coordinating scientific information management is structured in accordance with the following system life-cycle phases:

Standards Development

System development activities are most productive when they occur in an environment where information is easily shared, and new systems are built upon standard technologies and a common architectural design. ORD will benefit from policies, procedures, and standards that will allow the integration of future systems and maintain the quality and integrity of its published data. ORD should also be prepared to take advantage of new technologies and directions in science that can enhance the productivity of the bench scientist and allow users to access, use, and understand scientific information.

Requirements and Analysis

The first phase of the system development life cycle is the identification, documentation, and prioritization of requirements. Information is solicited from the appropriate users through joint requirements sessions, and current and future processes are examined. Business Process Re-engineering (BPR) may be required to make processes more efficient. Organizational and functional boundaries are examined to identify common requirements and shared data. Current systems are inventoried, and a gap analysis is performed to determine where current systems are inadequate.

System Design

System Design begins once the requirements have been identified and defined. When the requirements of a given system are not being addressed, or when the current system is leveraging obsolete technology, a new development effort is undertaken. A data repository may be designed using CASE tools to represent a logical view of the data. In addition, a user interface may be quickly developed using proposed technology to provide users with a prototype of the future system. Third-party software packages are often evaluated as part of this process.

Development and Implementation

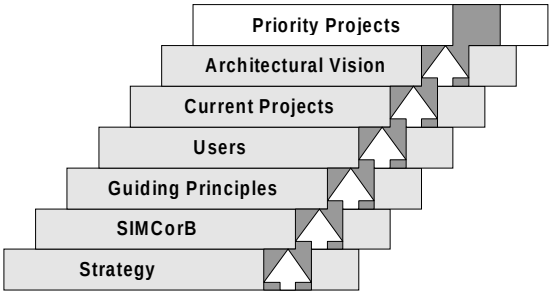
This phase involves the construction of the system in accordance with the prescribed design and the installation of the system in the production environment. This phase also includes the incorporation of necessary data into the new system and the training of users and system support staff prior to full system operation. In a RAD environment, the initial prototype may be enhanced using a continual, iterative process until it successfully meets the established requirements. Finally, the system is upgraded with enhanced security, online help, field validation and then migrated to a production environment.

Operations

This phase provides the full capabilities of the system to the users and ensures adequate, ongoing maintenance support for the system. It includes system modifications and enhancements, periodic formal evaluations of system, and the ultimate termination and archiving of the system at the end of its useful life.

Each of these phases addresses a unique dimension of the system development process, and each involves individuals with different organizational responsibilities and technical skills. For these reasons, the ORD organization for coordinating scientific information management is structured in accordance with the system development life cycle.

8 Priority Projects and Long-term Vision



Previous sections established a firm basis of information on ORD's strategic vision for scientific information management, user requirements, ORD laboratory and office responsibilities, and SIMCorB's role in achieving ORD's vision. This section describes several targeted initiatives that begin working toward realization of the ORD strategic SIM vision.

The earlier sections of this Plan provide a framework for transition from high-level policy goals and current practice to a rational future vision for scientific information management in ORD. This section outlines, at an intermediate level of generality, a series of implementation actions designed to systematically move toward realization of the SIM strategic vision. As the preceding discussion demonstrates, the scope of scientific information management is extremely broad, and the issues are complex. Not all major issues can be addressed at the same time. The projects described in this section represent initial priorities for implementation action; and although they do not comprise a comprehensive program for scientific IM improvements, they do provide essential first steps.

Project Management Approach

Following designation of operating project managers, detailed project management plans will be developed for each area of activity. Some current projects may be subdivided into several separate projects based upon differing time lines, resource availability, or technical requirements. The project management plans will address specific resource requirements, linkages with existing project activity and investments, timing, and performance-based milestones linked to intended outcomes.

As the details of these projects are developed, SIMCorB will develop an overall business plan that focuses on timing and source of resources, distribution of activity across ORD organizations, and linkages to the overall ORD and Agency IT/IM investment strategy.

This Plan is viewed as a living document, subject to operational revision over time within the context of the strategic vision and management values discussed in earlier sections. For example, additional SIM proof-of-concept projects related to ecological risk assessments, risk management technology assessment, or health risk assessment may be initiated as programmatic interest; organizational capabilities emerge, and project sponsors are identified.

The projects described in this section address a broad but selective range of SIM issues, all of which are important and should be addressed simultaneously. The overall pace of accomplishment will be governed in large measure by the aggregate level of resources committed to these projects. The priority projects address these four dimensions essential to improving scientific IM within ORD:

- People (ORD scientists, managers, and executives)
- Framework (comprehensive scientific IM vision and system architecture)
- Tools (improved policies, standards, procedures, analytical capability, and support systems)
- Partnerships (outreach and two-way collaboration across EPA, federal agencies, and the scientific community)

In addition to initiating and managing the listed projects, an early responsibility of each of the SIMCorB standing Sub-Groups will be to conduct a systematic review of policies, standards, procedures, and operations in their respective areas of responsibility. Thereafter they will advise SIMCorB of gaps in current scientific IM capabilities, significant vulnerabilities, and recommendations for future priority actions. In turn, SIMCorB will consolidate and balance identified needs and propose additional implementation actions for consideration by ORD management.

Priority Project Listing

The projects identified to date may be grouped and summarized in four categories:

1. Organization for Scientific Information Management

- Education and Training for Scientific Information Management Across ORD
- Enhanced Capacity for SIM Coordination in ORD Laboratories and Centers
- Fully Develop SIMCorB Coordination and Advisory Capability

2. Scientific IM Policies, Procedures and Standards

- Develop Data Administration and Quality Assurance Standards for Scientific IM
- User Requirements and Policies/Standards Development

3. Outreach Activities for Scientific Information Management

- Development of a SIMCorB Web Site to Promote Awareness of ORD SIM Activities
- Development of a Strategy to Promote Partnerships with Other Organizations

- Promote Awareness of SIMCorB and Scientific Information Activities Planned for ORD

4. Development of Common System Components for Scientific Information Management

- Implement Architecture for the Scientific Information Management System (SIMS) (This project includes data objects, metadata, scientific tools, and integrated ORD systems)
- Implement Architecture to Support Long-term Scientific Data Archives
- Implement Architecture to Support SIMS Integration With External Systems
- Evaluate New and Emerging Technologies for Scientific IM
- SIMS Proof of Concept Project: Information Management for Endocrine Disrupters
- SIMS Proof of Concept Project: Integration of UV-B Data Into SIMS
- Documentation of Priority ORD Data Sets and Analytical Tools

The following summary-level descriptions outline for each of the listed projects the current status of development; scope, timing, expected outcomes of future activities; related current scientific IM projects; and preliminary estimates of ORD staff and contractual resources required to achieve the planned objectives.

The resource estimates contained in the project descriptions are intended to provide an "order of magnitude" statement of projected requirements for strategic planning purposes. As noted, detailed resource plans will be developed for each project that consider current investments, opportunities for redirection of current or planned actions, and marginal requirements for new resources.

Long-term Vision for Scientific Information Management

Over the next five years, many scientific IM issues will be identified and solutions adopted through the SIMCorB process. These efforts should be guided by the realization that ever increasing amounts of environmental data and information will be generated by other Federal agencies and other national and international scientific organizations. Effective linkage of these information resources to environmental problem solving will require a consistent national (international) system for documentation and classification of scientific information. Successful implementation of the UV-B and Endocrine Disrupter projects described in this Plan will produce documentation of impediments to realizing that vision, as well as potential solutions. EPA and SIMCorB should provide leadership in seeking the creation of this type of national/international system.

8.1 Activities Supporting the Organization of ORD Scientific Information Management

8.1.1 Education and Training for Scientific Information Management Across ORD

A critical initial component of the ORD scientific information management strategy is the development of common understanding and shared views on the concepts, vocabulary, and current activities in scientific IM and the policy significance of scientific information management in accomplishing the EPA mission. This project will develop a design, curriculum specifications, training modules, and delivery process for education and training on scientific information management concepts, goals, and issues for:

1. ORD/EPA scientific personnel involved in generation of scientific information
2. ORD/EPA scientific personnel involved in the utilization of scientific information
3. ORD/EPA executives and science program managers

Current Status: Scientific information management concepts and issues are not well understood across the ORD community. Established project-focused science methods and documentation are not adequate to support extensive secondary use of scientific data or the demands of integrated assessment efforts. Inadequacies in scientific information management can impair the defensibility of ORD scientific information and conclusions based thereon.

Future Activity: Develop and implement a systematic program of training and education of scientific information management concepts, tools, and policy implications.

Expected Outcomes: Shared understanding of SIM concerns and voluntary employment of SIM standards, procedures and concepts in planning, and execution of ORD science activities.

Timing: Complete development and initiate implementation in FY99.

Project Manager: Thomas Lawless, National Exposure Research Laboratory

Resources:

ORD Staff:

Contract Work Years: FY99: 3.0; FY00: 2.0; FY01: 2.0; FY02: 2.0; FY03: 2.0

Related Current Projects:

6.4.5.5 Survey of ORD Data Management Systems
6.4.3.9 Information Tools Technology Training

8.1.2 Enhance Capacity for Scientific Information Management Coordination in ORD Laboratories and Centers

Achievement of a systematic approach to scientific IM across ORD will entail an on-site capability in each ORD operating organization to facilitate understanding of scientific IM concepts among ORD staff, support implementation of procedures and standards, and assure participation in ORD-wide IM development efforts. The character and scope of responsibilities, resources, and relationships required for this capability should be evaluated and options developed for consideration of ORD line management.

Current Status: Variable capacity for scientific information management coordination across ORD operating organizations.

Future Activity: Develop a requirements analysis and options paper on creating on-site scientific IM coordination and mentoring capability.

Expected Outcomes: Distributed institutional capability for scientific information management coordination and continuing education.

Timing: Complete by mid FY99.
Project Manager: To be determined.

Resources:
ORD Staff: Normal SIMCorB operations
Contract Work Years: FY99: 0.2.

Related Projects:
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.9 Information Tools Technology Training
6.4.3.10 Scientific Systems Requirements Analysis
6.4.5.5 Survey of ORD Data Management Systems
6.4.6.2 Management and Distribution of ORD Scientific and Technical Publications

8.1.3 Fully Develop SIMCorB Coordination and Advisory Capability

SIMCorB, the standing Sub-Groups, and the SIMCorB Executive Advisory Committee, in conjunction with ORD offices and staff, will be instrumental in the implementation of new scientific information management procedures and standards within ORD. SIMCorB operating procedures, roles, and responsibilities need to be clearly defined. An annual calendar needs to be established for SIMCorB meetings, document reviews, and recommendations to ORD management on policies and priorities. Leadership for standing Sub-Groups needs to be established, with clear support from ORD line management. In turn, procedures are required for securing participation of key ORD staff in Sub-Group operations with associated management support.

Current Status: The Charter for SIMCorB and SEAC have been adopted by ORD and Chairs of the Standing Sub-Groups designated.

Future Activity: Establishment of operating procedures, consultative processes, and a schedule of budgetary and programmatic reviews and recommendations as appropriate to the ORD Science, Management, and Executive Councils. Establishment of full membership of each Sub-Group and assessment of policies, procedures, and standards within the purview of each Sub-Group.

Expected Outcomes: Continuing capability for leadership and management of scientific information management requirements.

Timing: Complete organizational activity in FY99; continuing operations thereafter.

Project Manager: Robert Shepanek, NCEA

Resources:

ORD Staff: Continuing Sub-Group leadership commitment of 25% of work year for each Chair; participation of ORD staff in Sub-Group activities and management of projects.

Contract Work Years: Continuing support for SIMCorB, SEAC, and Sub-Group activities: FY99: 2.0; FY00: 2.0; FY01: 2.0; FY02: 2.0; FY03: 2.0

Related Projects:

- 6.4.1.1 Environmental Monitoring and Methods Committee (EMMC)
- 6.4.1.3 Supercomputing Executive Committee (SEC)
- 6.4.2.2 OMIS Expansion-Science IM Link
- 6.4.3.11 Enhanced Public Access Task Force—Environmental Monitoring for Public Access and Community Trading (EMPACT)
- 6.4.5.2 Human Exposure Information Management
- 6.4.5.5 Survey of ORD Data Management Systems

8.2 Projects Supporting Scientific Information Management Policies, Procedures, and Standards

8.2.1 Develop Data Administration and Quality Assurance Standards for Scientific IM

Enhanced scientific information management within ORD will require that a full range of data administration and quality assurance policies, procedures, standards, and oversight infrastructure have been established and are consistently implemented. These include:

- Develop records management and archival procedures and processes for scientific IM
- Develop metadata standards and procedures for scientific IM
- Develop data element naming standards for scientific IM
- Determine taxonomic, chemical, and methods standards for scientific IM
- Determine project documentation and data administration standards for scientific IM
- Develop risk assessment for scientific systems and security planning requirements
- Determine configuration management policies, procedures, and standards
- Create policies, procedures, and standards for the OMIS/QA tracking system

The Data Administration and Quality Assurance Sub-Group, in conjunction with line management and other SIMCorB Sub-Groups, will provide leadership in addressing these issues.

Current Status: Existing QA requirements cover documentation standards and requirements for some extramural IM issues, and ORD intramural science activities will be covered by forthcoming guidance. Standards for "data element naming" have been developed by OIRM/Enterprise Information Management Division (EIMD), and collaboration is underway with the Environmental Data Registry project on establishing common definitions for scientific data elements. Taxonomic, chemical, and methods standards have been developed both in interagency forums and in relation to specific EPA data bases; SIMCorB will coordinate with those efforts.

Future Activity: During FY98, policies and standards for archival and release of ORD scientific data and information will be completed and submitted for management review. Collaboration between ORMA and SIMCorB will begin on performing risk assessments for scientific systems and development of security planning requirements if warranted. Development of configuration management policies, procedures, and standards will be pursued in coordination with Advanced Technology's Sub-Group evaluation of tools. Recommendations for policies, procedures, standards, and oversight infrastructure for use in the OMIS/QA tracking system will be developed. An issues analysis and options paper will

be developed focusing on establishing consistent IM documentation standards for ORD-sponsored scientific activity under assistance agreements and interagency agreements.

Expected Outcomes:

Timing: Completion of five analysis projects by the end of FY99.
Project Manager: Linda Kirkland, NCERQA - Quality Assurance Division

Resources:

ORD Staff: To be determined.
Contract Work Years: FY99: 2.0; FY00: 1.0; FY01: 1.0; FY02: 1.0; FY03: 1.0.

Related Projects:

6.4.1.2 Models 2000
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.3 EIMS Data Types and Source Project
6.4.4.1 Scientific IM Links to ORD Competitive Research Grants—Science to Achieve Results (STAR)
6.4.4.2 EIMS Data Quality
6.4.5.1 Electronic Notebook Pilot Project
6.4.5.8 North Atlantic Research Strategy for Tropospheric Ozone (NARSTO)

8.2.2 User Requirements and Policies/Standards Development

ORD's information management resources need to be responsive to the needs of the organization's mission, individual research programs and projects, and to changing technologies. As the group that oversees the development of information management resources within ORD, SIMCorB requires development of mechanisms to insure that the requirements and expectations of ORD staff are known and understood. Those requirements will be met through the development of various systems and development of policies and standards necessary to support the stated requirements.

The objectives of this project are:

- Identify and implement mechanisms to identify and describe the information management requirements of ORD staff and management.
- Translate ORD scientist and manager requirements into information technology requirements and appropriate system development efforts.
- Develop guidance for selection of software, hardware, database design, and configuration design to achieve long-term interoperability within ORD.
- Develop policies and standards to support the information management requirements articulated by ORD staff.

Efforts to meet these objectives have commenced and will be more formally implemented as part of this project. Efforts will continue throughout the five-year period of this implementation coordination plan to ensure that information management activities are responsive to the evolving needs of ORD users.

Current Status: Informal steps were taken during the compilation of the ORD IRM Strategic Plan and this implementation coordination plan to identify and summarize ORD information management requirements. These requirements are reflected in the overall vision and specific projects outlined in this plan. Additional requirements have been collected from ORD scientists and managers as part of the system development and evaluation efforts for EIMS. Some EIMS requirements will be reflected in the revised EIMS database and application that will be released in July; additional requirements will be reflected in the fall release of EIMS. The Requirements Definition and Planning (REDAP) Sub-Group of SIMCorB coordinated an initial workshop involving representatives from various ORD data systems to identify common requirements.

Future Activity: The REDAP Sub-Group of SIMCorB will identify ORD requirements for information management through the implementation of survey tools, workshops, and focus groups. These requirements will be described and documented using a combination of informal reports, formal requirement documents, and CASE tools. In response to user defined

requirements, the REDAP Sub-Group will coordinate development and review of policies and standards that support the user requirements and the overall vision articulated in this plan.

Expected Outcomes: Requirements defined by this project will be used to form a more detailed vision for integrating information management resources within ORD. These requirements will also become drivers for other SIMCorB sponsored information management projects, such as the projects to implement the architecture for the ORD Scientific Information Management System (SIMS).

Timing: Efforts supporting this project have commenced and will continue throughout the five-year period represented by this implementation coordination plan.

Project Managers: Gary Collins, NERL-Cincinnati; and Jeffrey Frithsen, NCEA-Washington

Resources:

ORD Staff: Involvement of REDAP Sub-Group members with a commitment of 20% time per year for the five-year period of this plan. Periodic involvement of ORD staff involved with the development and/or management of ORD data systems.

Contract Work Years: FY99: 4.0; FY00: 2.0; FY01: 2.0; FY02: 2.0; FY03: 2.0.

Related Projects:

6.4.2.1 Year 2000 Project
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.5 Collaborative Work Process Infrastructure
6.4.3.7 Integrated Risk Information System (IRIS) Modernization
6.4.3.10 Scientific Systems Requirements Analysis
6.4.5.2 Human Exposure Information Management
6.4.5.3 THERDBASE
6.4.5.4 Consolidated Human Activity Database (CHAD)
6.4.5.5 Survey of ORD Data Management Systems
6.4.5.6 Capture of Data from Analytical Instruments
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 NERL Environmental Information Management System
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.7.3 Environmental Monitoring and Assessment Program (EMAP)

8.3 Projects Supporting Outreach Activities for Scientific Information Management

8.3.1 Development of a SIMCorB Web Site to Promote Awareness of ORD SIM Activities

Successful implementation of scientific information management efforts within ORD will require the support of the nationwide network of ORD staff, as well as program and data managers across EPA and other science organizations. Creation and management of a SIMCorB web site will be a useful technique to help promote awareness and understanding of the broad array of scientific IM initiatives.

Current Status: An internal SIMCorB homepage has been created. Each SIMCorB Sub-Group Chair will be responsible for a linked page related to the activities of their Sub-Group.

Future Activity: This effort will be coordinated with other outreach activities and will evolve with the program. The Sub-Group pages will be completed and updated regularly.

Expected Outcomes: As EIMS evolves into SIMS, the public will probably utilize that application for ORD scientific data, while the homepage may function as a communication mechanism primarily for SIMCorB management.

Timing: Efforts supporting this task are underway and will continue throughout the five-year period of this plan.

Project Director: Deric Teasley, NCEA

Resources:

ORD Staff: Involvement of the webmaster provided by NCEA will be 0.1 FTE. The various Sub-Group members will be involved periodically in updating the homepage for their Sub-Group.

Contract Work Years: FY99: 0.1; FY00: 0.1; FY01: 0.1; FY02: 0.1; FY03: 0.1.

Related Projects:

This task will be coordinated with the Outreach effort listed under Section 4.6.

- 6.4.1.1 Environmental Monitoring and Methods Committee (EMMC)
- 6.4.3.11 Enhance Public Access Task Force—Environmental Monitoring for Public Access and Community Trading (EMPACT)
- 6.4.4.1 Scientific IM Links to ORD Competitive Research Grants—Science to Achieve Results (STAR)
- 6.4.5.2 Human Exposure Information Management
- 6.4.6.2 Management and Distribution of ORD Scientific and Technical Publications

8.3.2 Development of a Strategy to Promote Partnerships with Other Organizations

Meeting the information requirements of the risk assessment/risk management process will increasingly entail access to data and information resources of other national and international scientific organizations. Understanding of the data resources and scientific IM capabilities of such organizations and establishment of working partnerships (individually or through interagency arrangements such as the Commerce, Energy, NASA, NLM, Defense Information Committee [CENDI]) can help achieve ORD's science objectives. Scientific information management efforts within ORD may be publicized through partnerships with other organizations, including REI and CENDI. Two-way information exchange and collaboration can enhance achievement of the EPA mission.

Current Status: There exists an extensive series of networks for project-level scientific collaboration and communication among ORD scientists and scientists of other organizations; however, system-level scientific IM seldom receives significant attention. A more targeted communications network focusing on scientific IM issues is emerging across the Federal government, as well as across other national and international science organizations. To date, ORD and EPA have not extensively capitalized on these emerging networks and the system-level, two-way, and multi-lateral collaboration opportunities associated therewith.

Future Activities: Development and implementation of a plan for collaborative partnerships and systematic involvement in national and international scientific IM networks for SIMCorB members and other ORD/EPA scientists and managers as appropriate. This will include formal membership and active participation in CENDI.

Expected Outcomes: Significantly improved systems for sharing of scientific data and information to support environmental and health risk assessment/risk management decisions

Timing: Assessment of partnership opportunities and plan development in FY98; implementation in FY99 and continuing.

Project Manager: Robert Shepanek, NCEA

Resources:

ORD Staff: To be determined.

Contract Work Years: FY99: 1.0; FY00: 1.0; FY01: 1.0; FY02: 1.0; FY03: 1.0.

Related Projects:

6.4.1.1 Environmental Monitoring and Methods Committee (EMMC)
6.4.5.8 North Atlantic Research Strategy for Tropospheric Ozone (NARSTO)
6.4.5.10 Multimedia Integrated Modeling System (MIMS)

8.3.3 Promote Awareness of SIMCorB and Scientific Information Activities Planned for ORD

ORD has been assigned scientific leadership within EPA. This responsibility includes both the direct conduct of scientific enterprise, as well as development and articulation of standards and processes that will promote the integrity and efficiency of environmental science across the Agency and in collaboration with external partners. SIMCorB was created in recognition of the unique requirements for scientific information management (SIM) in meeting the Agency's mission. SIMCorB's work towards the development of standards and systems to meet those requirements exemplifies how ORD is addressing the issue of scientific information. The full realization of the ORD vision for SIM will require the promotion and understanding of these requirements and systems among scientists, managers and policy officials across EPA.

Current Status: SIMCorB has been established to ensure that ORD's investments in scientific information resources are managed efficiently and effectively in support of the ORD vision of providing the scientific foundation to support EPA's mission of protecting human health and the environment. SIMCorB's Chair is tasked to provide overall executive leadership and ensure ORD-wide perspective and cross-organization integration. Each major organizational component of ORD provides a representative to SIMCorB who understands the scientific mission of their organization and the principles of scientific information management. The representatives are empowered to make technical and resource decisions on behalf of the organization with respect to the management of scientific information.

The early efforts of SIMCorB members focused on fostering team work and establishing relationships and responsibilities within the board and its sub-groups. One of the first deliverables developed by SIMCorB has been this Implementation Coordination Plan, which has itself proceeded through various peer and management reviews. The SIM strategy developed by ORD provides a vehicle to promote common understanding of the processes that provide the scientific information needed to meet EPA's mission responsibilities as well as processes to assure the efficient acquisition of appropriate IT/IM capabilities. However, awareness and understanding of the issues within ORD and among the EPA program and management offices is limited. Successfully moving from the strategic planning stage on to project implementation will require a concerted effort to present the SIM vision, strategy, and tactics developed by SIMCorB to the management, scientific, and administrative staff of ORD's labs and centers.

Future Activity: Develop appropriate briefing materials and handouts, arrange logistics, establish site visit schedule, and systematically conduct a series of briefings with relevant managers, scientists and policy officials at each individual ORD operations site and across the EPA program and management offices. Briefings would focus on the role of scientific information in meeting EPA's mission responsibilities and the project management issues and tactics for implementing specific projects in support of effective scientific information

management. Develop a Lotus Notes discussion database to facilitate two-way communication between SIMCorB and ORD staff and to provide a readily-available mechanism to collect comments and reactions. These outreach activities would provide a policy-level complement to the SIM education effort anticipated in project 8.1.1 in this Plan.

Expected Outcomes: Establish a common policy-level understanding of SIM issues and strategy across EPA program and management offices, and at each ORD site.

Timing: Develop briefing material and initiate briefings in FY 99; continue as needed in FY 00 and FY 01.

Project Manager: Mike Waters (??)

Resources:

ORD Staff:
Contract Work Years: FY 99:1.0; FY 00: 0.5; FY 01: 0.5

Related Current Projects:

6.4.1.2 Models 2000 Steering/Implementation Plan
6.4.1.3 Supercomputing Executive Committee (SEC)
6.4.2.2 OMIS Expansion- Science IM Link
6.4.3.12 EPA Risk Assessment Forum
6.4.4.1 Scientific IM Links to ORD Competitive Research Grants

8.4 Projects Supporting Development of Common System Components for Scientific Information Management

8.4.1 Implement Architecture for SIMS

This project will systematically implement key components of the SIMS architecture defined in Section 7. Priority will be assigned to: data object architecture, metadata architecture, open access to scientific tools, and ORD system integration.

Data Objects: Centralized storage and access to all ORD data objects leverages existing investments in research, promotes information sharing within the organization, and minimizes duplication of effort among researchers. The architecture should support all data object types, including but not limited to: projects, models, data sets, databases, spatial data sets, and documents.

Metadata: A repository of all ORD scientific metadata information can provide a central point of access for all scientific data within ORD. Such a configuration would ease access to metadata and data, promote information sharing, and reduce unwarranted redundancy in data development.

Scientific Tools: ORD can leverage existing scientific tools and minimize future investments by developing an architecture that allows multiple ORD systems to access and use scientific tools residing a centrally-available repository. Such a configuration would minimize duplication of scientific tools, promote broader access to a full range of scientific tools, and minimize version control issues.

Integrated ORD Systems: ORD information systems should be logically networked to allow users to access information and use functionalities in multiple systems. Such an architecture minimizes administrative overhead, enhances quality assurance, promotes information sharing within the organization, and leverages existing financial and time investments made in current systems, including OMIS, EIMS, and QA tracking.

Current Status: Efforts to develop a common architecture for the ORD information management environment have commenced. EIMS is developing an architecture that supports distributed searches of metadata information such as to realize the vision of providing an Agency inventory of environmental information. This vision was originally articulated in the IM strategic white paper (Shepanek, 1997) produced by NCEA. Through the EIMS distributed architecture, users will be able to access metadata about multiple objects (data sets, databases, projects, models, documents, multimedia products, etc.), and in some cases, obtain access to the objects themselves.

The Models 2000 effort is developing approaches for improving the interoperability of models used by the Agency. Platform independent software is being developed to link models to multiple databases, and integrate models such that the output of one model can be used to fulfill the input requirements for a second model.

Future Activity: Future activities of this project will focus on developing enhanced approaches to integrating metadata and data systems across the Agency, and with systems maintained by other organizations. These approaches will make it possible to implement for ORD scientists and managers, integrated desktop tools such as the geographic information module (GEOSIM) and the modeling module (MODSIM) of the ORD SIMI. Pilot projects will be completed to develop approaches for mining metadata from selected data systems. Standards will be developed to support the integrated architectural environment. Additional efforts planned for this project will be defined as they are identified.

Expected Outcomes: This project will provide the common architecture for the suite of tools included in the ORD SIMI and enable the interoperability of those tools.

Timing: Efforts supporting this project have commenced and will continue throughout the five-year period represented by this implementation coordination plan.

Project Managers: Joan Novak, NERL-Research Triangle Park; and Robert Shepanek, NCEA-Washington

Resources:

ORD Staff: This project will involve participation from members of the Advanced Technology Evaluation and Modeling (ATEM) Sub-Group, the Requirements Definition and Planning Sub-Group, and the Systems Engineering and Operations Sub-Group of SIMCorB. A commitment of 25% time per year for the five-year period of this plan will be required from ATEM members. Periodic involvement of ORD staff involved with the development and/or management of ORD data systems.

Contract Work Years: FY99: 5.0; FY00: 2.5; FY01: 1.5; FY02: 1.5; FY03: 1.5.

Related Projects:

Data Objects:

6.4.3.1 Environmental Information Management System
6.4.3.2 National Health and Nutrition Examination Survey (NHANES)
6.4.3.3 EIMS Data Types and Source Project
6.4.3.7 Integrated Risk Information System (IRIS) Modernization
6.4.5.3 Total Human Exposure Risk Database and Advanced Simulation Environment (THERDBASE)
6.4.5.4 Consolidated Human Activity Database (CHAD)
6.4.5.7 Information Management for NERL Endocrine Disruptor Research

6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.6.2 Management and Distribution of ORD Scientific and Technical Documents
6.4.7.3 Environmental Monitoring and Assessment Program (EMAP)
6.4.5.14 National Human Exposure Assessment Survey (NHEXAS)

Metadata:
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.1 Environmental Management Information System (EIMS)
6.4.4.1 Scientific IM Links to ORD Competitive Research Grants—Science to Achieve Results (STAR)
6.4.5.1 Electronic Notebook Pilot Project
6.4.5.2 Human Exposure Information Management
6.4.5.7 Information Management for NERL Endocrine Disruptor Research
6.4.5.8 North Atlantic Research Strategy for Tropospheric Ozone (NARSTO)
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 NERL Environmental Information Management System
6.4.5.12 Environmental Monitoring and Assessment Program (EMAP)

Scientific Tools:
6.4.1.1 Environmental Monitoring and Modeling Committee (EMMC)
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.10 Scientific Systems Requirements Analysis
6.4.5.2 Human Exposure Information Management
6.4.5.5 Survey of ORD Information Management Systems
6.4.5.7 Information Management for NERL Endocrine Disruptor Research
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 NERL Environmental Information Management System
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.7.3 Environmental Monitoring and Assessment Program

Integrated ORD Systems:
6.4.2.2 OMIS Expansion-Science IM Link
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.3 EIMS Data Types and Source Project
6.4.3.4 Bibliographic Software Assessment
6.4.3.7 Integrated Risk Information System (IRIS) Modernization
6.4.4.2 EIMS Data Quality
6.4.5.2 Human Exposure Information Management
6.4.5.3 Total Human Exposure Risk Database and Advanced Simulation Environment (THERDBASE)
6.4.5.4 Consolidated Human Activity Database (CHAD)

6.4.5.7 Information Management for NERL Endocrine Disruptor Research
6.4.5.8 North Atlantic Research Strategy for Tropospheric Ozone (NARSTO)
6.4.5.9 UVB Pilot Project Data
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 NERL Environmental Information Management System
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.6.2 Management and Distribution of ORD Scientific and Technical Documents
6.4.6.3 Alternative Treatment Technology Information Center (ATTIC)
6.4.7.1 National Health and Nutrition Examination Survey (NHANES)
6.4.7.3 Environmental Monitoring and Assessment Program (EMAP)
6.4.5.14 National Human Exposure Assessment Survey (NHEXAS)

8.4.2 Implement Architecture to Support Long-term Scientific Data Archives

A consistent ORD and EPA policy framework for records management, including disposal by archiving, will provide the necessary context for archive system development and implementation. ORD personnel will develop an ORD-wide records management plan coordinated with roles and responsibilities of users for managing scientific data base objects in coordination with the EPA Electronic Records Specialist. Evaluation of Agency proposals for records management software will be carried out to see if they can adequately handle linkage between records. These activities will provide information on the scope and nature of archival requirements for ORD scientific databases. The subsequent archive architectural plan would allow multiple connected ORD systems to access archived scientific information of any type. Archived data should include Models, Datasets and Databases (including model runs), Software Systems and Tools, Documents and Methods, EIMS Master Metadata, annual OMIS information, ORD Research Products, and Web Pages.

Current Status: Current EPA record retention policies do not adequately address criteria and management processes for retention and archiving of ORD intramural research data and information. Information technology and standards for long term storage and continuously usable public access to ORD scientific information have not been fully evaluated. Techniques for linkages to laboratory notebooks and project files have not been developed. EPA's National Records Management Program has convened a Records Management Application (RMA) workgroup that includes ORD representation, and two RMAs are being pilot tested in Region 5.

Future Activity: This project will include convening of a cross-ORD working group to develop recommended criteria and processes for record management determinations on research data "of continuing research value" (thus subject to long-term archiving requirements), an ORD-wide record management plan, and record management software evaluation for scientific database objects. In addition, the project will develop options for SIMCorB consideration dealing with search engines, and security and data base design to support usable access to stored or archived research data and its descriptive metadata. ORD/SIMCorB representatives will participate in the Agency-wide work groups to assure effective linkages between ORD and EPA strategies as well as ensuring that Agency policies consider the requirements of scientific information management.

Expected Outcomes: Establishment of a policy, an ORD-wide record management plan, and IM structure to support usable public and scientific access to stored scientific information developed by ORD.

Timing: Policy and record management plan development will begin in FY98, with recommendations to ORD management in mid FY99. System assessments focused on record

management software will begin early in FY99, with implementation of selected options in FY99 and FY00.

Project Manager: Brenda Culpepper, NERL; and Linda Kirkland, NCERQA-Quality Assurance Division

Resources:

ORD Staff: To be determined.

Contract Work Years: FY99: 2.0; FY00: 0.5; FY01: 0.5; FY02: 0.5; FY03: 0.5.

Related Projects:

6.5.4.1.1 Environmental Monitoring and Methods Committee (EMMC)
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.4 Bibliographic Software Assessment
6.4.3.11 Enhanced Public Access Task Force—Environmental Monitoring for Public Access and Community Trading (EMPACT)
6.4.5.2 Human Exposure Information Management
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 NERL Environmental Information Management System
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.6.2 Management and Distribution of ORD Scientific and Technical Documents
6.4.7.3 Environmental Monitoring and Assessment Program (EMAP)

8.4.3 Implement Architecture to Support SIMS Integration with External Systems

ORD research and assessment activities are becoming more complex involving multiple investigators representing multiple disciplines, and working at multiple spatial and temporal scales. Increasingly, ORD scientists and associated partners rely on non-ORD sources of data and information to complete these assessments. To be effective, ORD information management resources must provide efficient and easy access to information about other sources of data, provide mechanisms for easy access to those data, and develop approaches for integrating data from multiple sources of data into a consistent organization to enable integrated assessments.

The objectives of this project are:

- integrate ORD environmental resource inventory efforts with other inventory efforts within the Agency.
- identify opportunities for applying and developing standards and approaches that will support finding and using data located outside of the Agency.
- develop approaches to facilitate integration of data collected and maintained by different organizations and users.

This project represents an extension of current ORD and Agency activities; however, specific directions and priorities for this project will be partially driven by the requirements identified as part of the project defined in Section 8.2.2.

Current Status: The existing ORD EIMS is being used to integrate metadata and data from an ORD research project (ReVA), an ORD research center (NCEA), a program office (Office of Water Surf Program and the Index of Watershed Indicators [IWI] database), and an EPA Region (Region 10). Thus, EIMS is already providing links to non-ORD data systems within the Agency. Further, EIMS has been designed to be consistent with the National Aeronautics and Space Administration (NASA) data interchange format (DIF) standard, and the FGDC content standard for spatial metadata, thus supporting interagency consistency for scientific metadata development. The EIMS team has commenced discussions with the development team for EDR concerning the integration of the EIMS and EDR. Planning has commenced to have EIMS become a node as part of the NSDI, enabling distributed searches of spatial data.

Future Activity: A plan will be produced to outline and set milestones for the integration of EIMS and EDR. The team for this project will also work with the SIMCorB REDAP Sub-Group to identify other data bases that should be linked or integrated with the ORD SIMS to support assessment and research activities. It is envisioned that a plan and schedule will be developed to describe integration approaches and time lines. The team for this project will also work with the SIMCorB Advanced Technology Evaluation and Modeling Sub-Group to

identify new and emerging information technologies that will support more efficient integration of disparate databases.

Expected Outcomes: This project will result in enhanced integration of databases developed within the Agency, and more efficient approaches to access and use data in systems maintained by other organizations. Developing approaches that use information technologies to more efficiently integrate data will enable ORD scientists to spend more time on problem formulation and data analysis, and less time on merging data from multiple sources.

Timing: Efforts supporting this project have commenced and will continue throughout the five-year period represented by this implementation coordination plan.

Project Director: Jeffrey Frithsen, NCEA-Washington

Resources:

ORD Staff: Involvement of the REDAP and ATEM Sub-Group members with a commitment of 20% time per year for the five-year period of this plan. Periodic involvement of ORD staff involved with the development and/or management of ORD data systems.

Contract Work Years: FY99: 3.0; FY00: 2.5; FY01: 2.0; FY02: 2.0; FY03: 2.0.

Related Projects:

6.4.1.1 Environmental Monitoring and Methods Committee (EMMC)
6.4.3.1 Environmental Information Management System (EIMS)
6.4.3.4 Bibliographic Software Assessment
6.4.3.11 Enhance Public Access Task Force—Environmental Monitoring for Public Access and Community Trading (EMPACT)
6.4.5.2 Human Exposure Information Management
6.4.5.10 Multimedia Integrated Modeling System (MIMS)
6.4.5.11 Environmental Monitoring Information Management System
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.6.2 Management and Distribution of ORD Scientific and Technical Documents
6.4.7.3 Environmental Monitoring and Assessment Program (EMAP)

8.4.4 Evaluate New and Emerging Technologies for Scientific IM

Constant evaluation of new technology will ensure that ORD systems development efforts are completed using the most capable software and hardware available. The Advanced Technology Evaluation and Modeling Sub-Group will continually assess new and emerging technologies as they are released.

Current Status: Existing technology assessments activities related to scientific IM tend to be sporadic, fragmented, and lacking in system focus.

Future Activities: Establish a systematic process for SIM technology assessment organized in reference to the SIMS architectural framework. Establish links with public and private sector groups concerned with SIM technology evaluation.

Expected Outcomes: Timely identification and evaluation of new technologies to support the ORD mission.

Timing: Project initiation in FY 98; continuing activity thru FY 01.

Project Manager: Joan Novak, NERL

Resources:

ORD Staff: To be determined.

Contract Work Years: FY99: 0.5; FY00: 0.5; FY01: 0.5; FY02: 0.5; FY03: 0.5.

Related Projects:

6.4.5.10 Multimedia Integrated Modeling System (MIMS)

8.4.5 SIMS Proof of Concept Project - Information Management for Endocrine Disruptors

Full mobilization of the scientific resources of ORD in the three National Laboratories and two National Centers will require a sophisticated information management system to support data and information sharing among widely dispersed researchers working on different dimensions of the science problem. It must also support the archive and recovery of critical information over time.

The following demonstration project, as currently conceived, will involve development of distributed databases on exposure, health and ecological effects, and risk assessment for endocrine disrupting chemicals and associated metadata. The project will provide a test-bed for many emerging IM concepts and the integrated scientific information management system (SIMS), while simultaneously solving a key science problem for ORD and the Program Offices.

A Cross-Organizational Example of the System Development Approach

The first "proof of concept" will be the design, development of architecture, and implementation of a metadata management capability for EPA/ORD including endocrine disruptor data and documents assembled by NHEERL, NERL, and NCEA, as well as the National Center for Toxicological Research (FDA) as described below. The effort will link to the EPA Endocrine Disruptors Web site <www.epa.gov/endocrine> which is used by the federal government, states, academia, the private sector, and environmental organizations internationally to determine the current status of research on the topic.

This activity will parallel the creation of a global inventory of endocrine disruptor research projects by NHEERL to be maintained ultimately by the European Commission DG XII Joint Research Centre in Ispra, Italy. The global inventory of endocrine disruptor research will be based on the format of the Committee on Environment and Natural Resources (CENR) inventory of U.S. federally funded research which includes: title and short abstract; funding agency and contact; focal areas of human health, ecology, and exposure; research categories of methods, models, and measurements; research subcategories of mixtures, biomarkers, basic research, sentinel species, and others; and keywords for endpoints, systems, and agents.

As a result of SIMCorB efforts, in addition to its present attributes, the EPA Endocrine Disruptor Web site at <www.epa.gov/endocrine> will include a metadata management and data analysis system to support integrated health and ecological risk assessment on endocrine disrupting chemicals. The Endocrine Disruptor Data Directory will be an Oracle database used to manage and track data sets. The Directory will be queried using various keywords to find data sets of interest. The Endocrine Disruptor Data Catalog will prescribe a format for metadata. Every data set will be accompanied by a metadata file. The Endocrine Disruptor

internal Web site in RTP will contain data sets, metadata files, publications, maps, and program information and will be used for development, testing, and internal use.

Endocrine Disruptor data sources within ORD are the NHEERL and NERL Divisions involved in the Endocrine Disruptor Research Program. Other data sources are external to ORD and include FDA and other organizations in this country and abroad. These organizations, as they become involved in the effort, will prepare summary data sets, metadata files, and publications to put on the Endocrine Disruptor Web site.

8.4.5.1 The Proposed Endocrine Disruptor Database (EDD)

Under an interagency collaborative effort led by the Office of Prevention Pesticides and Toxic Substances (OPPTS), an Endocrine Disruptor Database is currently being developed. Three software systems relevant to the EDD are presently available in EPA/OPPTS, FDA/NCTR, and EPA/ORD:

I. The prototype Endocrine Disruptor Priority Setting Database (EDPSDB). EDPSDB was developed to facilitate priority setting in endocrine disruptor screening. It contains records for ~87,000 chemicals with Chemical Abstracts Service Registry Numbers (CASRNs) from data sources related to the information categories and criteria described in the EDSTAC, April 3, 1998, draft report. The database was created using Molecular Design Limited Information Systems' Integrated Scientific Information System (ISIS/Base). It contains detailed structural information for each compound, as well as extensive data on statutory, exposure, and effects-related criteria. The system is noteworthy for its structural search capabilities and for its ability to support complex boolean queries on the associated compound information.

II. The Estrogen Knowledge Base (EKB). The EKB is maintained at the National Center for Toxicological Research. The EKB was developed to house detailed experiment results on endocrine disrupting compounds, and to serve as a basis for data exploration and the development of predictive models. The EKB database is implemented in the Oracle relational database using Java language applets to provide access through the World Wide Web. The EKB database can store experiment results for many different assays across multiple compounds and species. Each data point is linked to a bibliographic reference indicating its source, such as a journal article or laboratory report. The EKB stores detailed experiment results in a way that eases comparison and analysis, while providing easy access through web browsers such as Netscape Communicator and Microsoft Explorer.

III. Graphic Activity Profiles (GAP). GAP software used in the EPA/IARC Genetic Activity Profile Database provides a method of graphically characterizing a compound's activity as a function of dose in a compact, easily readable histogram. The format has already been adapted to developmental toxicity data (Kavlock, et al., Teratology, 43:159-185, 1991). In addition, GAP ISO software can identify chemicals that display similar biological activity as a function of dose. GAPs will be used to present an overview of the endocrine activity of chemicals that

have been evaluated using the battery of tests recommended by EDSTAC as well as to display other test results supported by detailed data listings. It is envisaged that the Endocrine Disruptor Profile will be the primary user interface for access to effects (and perhaps exposure) data on putative endocrine disrupting chemicals.

These three software systems will be combined into a single EDD. This will give researchers and regulators the ability to easily move between the overview of the data provided by the prototype EDPSDB and the detailed experiment data stored in the EKB, while providing the convenience of a graphic display of endocrine activity profiles. The common database platform will be Oracle. Oracle is the database standard both within EPA and within the FDA. The EKB is already implemented in Oracle. The prototype EDPSDB is currently in ISIS/Base, however the developers' intention was to eventually keep much of the data in Oracle. This combined Endocrine Disruptor Database will preserve the capabilities of the prototype EDPSDB, while adding the detailed experiment data structures and web accessibility of the EKB. It will also incorporate the algorithms, or possibly the code, of Graphic Activity Profiles.

Current Status: Discussions are currently under way regarding the combining of the three databases described above. Additional databases on exposure and ecological effects will be added. Once decisions are made regarding these fundamental components, detailed planning can begin on the design of the IM details that will fall to SIMCorB.

Future Activity: As SIMCorB undertakes system-level assignments and coordinates the design and development of data, function, and technology architectures, the SIMCorB Sub-Groups will follow the system development approach outlined earlier in this section.

Expected Outcomes: Development of distributed databases on exposure, health and ecological effects, and risk assessment for endocrine disrupting chemicals and associated metadata. Development of a sophisticated information management system to support data and information sharing among widely dispersed researchers, and to support the archive and recovery of critical information over time. In addition, as stated above, the project will provide a test-bed for many emerging IM concepts.

Timing: Estimation of timing in detail is premature at this point since the activities described above have just begun. However, timelines will clearly be driven by the requirements of Congressional legislation under the Food Quality Protection Act and the amendments to the Safe Drinking Water Act, which were passed in the summer of 1996. Both of these acts require that the Agency develop a screening and testing strategy for endocrine disruptors by August, 1998; implement screening and testing by August, 1999; and report progress to Congress by August, 2000.

Project Manager: Michael Waters, NHEERL

Resources:

ORD Staff: To be determined
Contract Work Years: FY99: 2.0; FY00: 3.0; FY01: 1.0; FY02: 1.0; FY03: 1.0.

Related Projects:

6.4.5.7 Information Management for NERL Endocrine Disruptor Research
6.4.7.5 Quantitative Structural Activity Relationship (QSAR) Modeling Systems
6.4.7.6 ECOTOX
6.4.7.7 EPA/ International Agency Research on Cancer (IARC) Graphic Activity Profile (GAP)
6.4.7.8 Information Management for NHEERL Endocrine Disruptor Research

8.4.6 Proof of Concept Project: Integration of UV-B Data into SIMS

The UV-B Data Project (5.4.5.9) is a current development project that features the integration of summary-level UV-B data for use in SIMS, and illustrates the benefits of an integrated information system within ORD. Prior to its inclusion within the context of SIMS, the UV-B data was accessible to a very small group of scientists, and in a limited form. With the goals of providing data accessibility to external scientists and the general public, as well as preserving data integrity and establishing an archival methodology, NERL initiated the project of integrating summary-level UV-B data into EIMS, the precursor to SIMS.

The original scope of the UV-B Data Project specified creation of an Oracle RDBMS and a web-enabled user interface providing data assessment capabilities that include queries, calculations, and graphs. This development effort is currently underway; the technical and functional specifications for the project led SIMCorB to target UV-B as a possible candidate to demonstrate scientific information integration into SIMS.

To the meet the Agency's mission of risk assessment and information management, the scope of the UV-B development effort has been significantly expanded to include integration with SIMS. Via SIMS, UV-B users will have a single point of access to other ORD data sets and databases, as well as metadata, software, models, analysis, and visualization tools. The UV-B/EIMS integration effort will be used to establish a precedent for all data integration with SIMS, and its interoperability with other distributed data, metadata, tools, and documentation.

Current Status: Analyze the available data to generate specifications for the database. Create logical and physical design specifications for the UV-B application.

Future Activity: Legacy data will be loaded into the new database and the mechanism will be developed to obtain daily reports from the field instruments. Raw data will be available the day after it is collected. An archival mechanism will be put in place as that capability is added to EIMS. The query capabilities of the relational database will be made available as the Assessment Query Interface (AQI) is developed under EIMS.

Expected Outcomes: The choices made in setting up the UV-B database will set up a methodology for dealing with future unique data sets. The requirement for the AQI will put development of this functionality on a high priority time line. UV-B data will be readily available to researchers around the world.

Timing: The Web page will be available in the first quarter of FY99. Integration of the EIMS capabilities for metadata will follow in the second quarter with the full AQI functional in the fourth quarter of FY99.

Resources:

ORD Staff: To be determined.

Contract Work Years: FY99: 2.0; FY00: 0.5; FY01: 0.2; FY02: 0.1; FY03: 0.1.

Related Projects:

5.4.5.7 Information Management for NERL Endocrine Disruptor Research

8.4.7 Documentation of Priority ORD Data Sets and Analytical Tools

As discussed throughout this SIM Implementation Coordination Plan, the nature of environmental science is changing. The analyses and assessments needed to address environmental problems require integrating multiple kinds of data at multiple scales. Many of these data will have been developed over varying time periods in seeking answers to diverse scientific questions, but which may be relevant to contemporary or future environmental assessment efforts. The effective secondary use of such information resources requires an ability to identify the existence, character and location of the data and the analytical tools used in the prior studies, as well as understanding of access paths and limitations imposed by the data manager.

Current Status: The traditional practice for scientific data management has revolved around the individual or small team of scientific investigators and the publication of scientific papers attributable to members of the research group. In general, data sets associated with traditional experimental research were small and data characteristics were described in appropriate detail within the body of the publications. Increasingly, however, contemporary environmental science relies upon multiple large data sets, complex analytical tools and models, and sophisticated processing routines in the conduct of risk assessment and risk management analyses. Traditional data management practices are inadequate to support this type of scientific effort.

The Scientific Information Management System (SIMS) envisioned in this Plan creates a management framework for organizing and providing access to essential information (meta data) on environmental information resources deemed critical to future scientific efforts. To date, a systematic process has not been established for identification of critical information resources and development of consistent meta data documentation. This has contributed to the inaccessibility and/or loss of some significant information resources, and the duplicate storage of other data.

Future Activity: Develop and implement a systematic process for:(1) evaluation of legacy (historical) information resources in regard to their potential utility in future environmental science; (2) for the information resources deemed to have long term significance, the consistent documentation of meta data for the data sets, analytical tools and unique information handling routines; and, (3) systematic entry of such meta data into SIMS.

Expected Outcomes: Establishment of ORD-wide criteria for classification of information resources with long term significance for application to both historical and contemporary science efforts, identification of critically important legacy information resources, and development of initial meta data entries into an inventory of environmental information resources accessible to all scientists, managers and policy officials.

Timing: Project will be initiated in FY 99 and continue through FY 03.

Project Manager: (to be defined)

Resources:

ORD Staff:
Contract Work Years: FY99: 1.0; FY00 : 1.5; FY 01: 1.5; FY02: 1.5; FY03: 1.5.

Related Current Projects:

6.4.3.1 Environmental Information Management System (EIMS)
6.4.5.9 UV-B Data Project
6.4.5.12 Regional Vulnerability Assessment (ReVA) IM Support
6.4.7.2 Environmental Monitoring and Assessment Program (EMAP)
6.4.7.4 ECOTOX

Appendix A. List of Acronyms

AQI	Assessment Query Interface
AQUIRE	Aquatic Toxicity Information Retrieval
ASTER	Assessment Tool for Ecological Risk
ATEM	Advanced Technology Evaluation and Modeling
ATTIC	Alternative Treatment Technology Information Center
BASINS	Better Assessment Science Integrating Point and Nonpoint Source
BPR	Business Process Reengineering
CASE	Computer Aided Software Engineering
CASRNs	Chemical Abstracts Service Registry Numbers
CDC	Centers for Disease Control
CEIS	Center for Environmental Information and Statistics
CENDI	Commerce, Energy, NASA, NLM, Defense Information Committee
CENR	Committee on Environment and Natural Resources
CHAD	Consolidated Human Activity Base
CIO	Chief Information Office
DIF	Data Interchange Format
DOCSIM	Document management component of SIMS
ECOTOX	Ecotoxicology Database Retrieval System
EDD	Endocrine Disruptor Database
EDPSDB	Endocrine Disruptor Priority Setting Database
EDR	Environmental Data Registry
EIMD	Enterprise Information Management Division
EIMS	Environmental Information Management System
EKB	Estrogen Knowledge Base
EMAP	Environmental Monitoring and Assessment Program
EMMC	Environmental Monitoring and Methods Committee
EMPACT	Environmental Monitoring for Public Access and Community Trading
FGDC	Federal Geographic Data Committee
FOIA	Freedom of Information Act
GAP	Graphic Activity Profile
GEOSIM	Geographic Information Module of SIMI
GIS	Geographic Information System
GPRA	Government Performance and Results Act
HSPF	Hydrologic Simulation Program Fortran
IARC	International Agency Research on Cancer
IM	Information Management
IRDS	Information Resources Development Staff
IRIS	Integrated Risk Information System
IRM	Information Resources Management

ISIS	Integrated Scientific Information Systems
ITMRA	Information Technology Management Reform Act
IWI	Index of Watershed Indicators
LABSIM	Laboratory information management component of SIMS
LIPS	Laboratory Implementation Planning System
MIMS	Multimedia Integrated Modeling System
MIS	Management Information System
MISS	Management Information System Staff
MODSIM	Modeling Component of SIMI
NADS	Network Analytical Data System
NARSTO	North Atlantic Research Strategy for Tropospheric Ozone
NASA	National Aeronautics and Space Administration
NBII	National Biological Information Infrastructure
NCEA	National Center for Environmental Assessment
NCERQA	National Center for Environmental Research and Quality Assurance
NERL	National Exposure Research Laboratory
NHANES	National Health and Nutrition Examination Survey
NHEERL	National Health and Environmental Effects Research Laboratory
NHEXAS	National Human Exposure Assessment Survey
NIH	National Institute of Health
NOAA	National Oceanic and Atmospheric Administration
NRMRL	National Risk Management Research Laboratory
NSDI	National Spatial Data Infrastructure
OIRM	Office of Information Resources Act
OLAP	Online Analytical Processing
OMIS	ORD Management Information System
OPPE	Office of Policy, Planning, and Evaluation
OPPTS	Office of Prevention Pesticides and Toxic Substances
ORD	Office of Research and Development
ORMA	Offices of Resources Management and Administration
OSP	Office of Science Policy
PALMSIM	Palm computer interface component of SIMS
PATRIOT	Pesticide Assessment Tool for Rating Investigations for Transport
PAVE	Package for Analysis and Visualization of Environmental Data
PHYTOTOX	Terrestrial Plant Toxicity Database
PRZM	Pesticide Root and Zone Model
QA	Quality Assurance
QA/QC	Quality Assurance/Quality Control
QSAR	Quantitative Structural Activity Relationship
RAD	Rapid Application Development
RADM	Regional Acid Deposition Model
RDBMS	Relational Database Management System

REDAP	Requirements Definition and Planning
REI	Reinventing Environmental Information
ReVA	Regional Vulnerability Assessment
RMA	Records Management Application
SAB	Science Advisory Board
SAS	Statistical Analysis Software
SDEP	State's Department of Environmental Protection
SDFW	SIMS Data Format Wizard
SEAC	SIMCorB Executive Advisory Committee
SEC	Supercomputing Executive Committee
SIM	Scientific Information Management
SIMCorB	Science Information Management Coordination Board
STAR	Science to Achieve Results
SIMI	Scientific Information Management Interface
SIMS	Scientific Information Management System
STATSIM	Statistical toolkit to SIMI
TERRETOX	Terrestrial Wildlife Toxicity Database
THERDBASE	Total Human Exposure Risk Database and Advanced Simulation Environment
UV-B	Ultraviolet-B
WAN	Wide Area Network

Appendix B. Strategic Plan for the Office of Research and Development: Information Management Component

United States
Environmental Protection
Agency

Research and
Development
(8101)

EPA/600/R-97/086
August 1997

Strategic Plan for the Office of Research and Development

Information Management Component

Printed on paper that contains at least 20% postconsumer fiber.

Contents

Glossary/List of Acronyms

architecture	In the context of an information management system, architecture means the overall logical and physical definition of the system.
browser	An electronic tool for reviewing information in a database. Also referred to as a "database browser."
catalog	A set of detailed documentation about a number of data sets that helps users determine which data sets may be useful for a particular application. A catalog provides much more detailed information than a directory.
CBEP	EPA's Community-Based Environmental Protection initiative.
CD-ROM	Literally, Compact Disc-Read-Only Memory. A type of compact disc that is used to store software, databases, and reusable information.
CENR	The White House Committee on Environment and Natural Resources.
CIO	Chief Information Officer; the person who has been officially designated as having the overall responsibility and authority for information management activities, policies, and direction in an organization.
client/server	Use of distributed "client" computer systems linked to a central "server" in order to share common software tools/applications and capability.
data set	A logical meaningful grouping or collection of similar or related data.
DCD	directory/catalog/dictionary.
dictionary	A set of basic information about individual components of a single data set. For example, a dictionary for a scientific data set would include format information and a short scientific description of the parameters or variables in the data set.
directory	A set of information about a large number of data and information products that informs users which products are available and generally what they contain.
GIS	Geographic Information System; a collection of computer hardware, software, and geographic data designed to capture, store, update, manipulate, analyze, and display geographically referenced data.
GPRA	The Government Performance and Results Act.
hypertext	A form of electronic document that contains links to other electronic documents. Typically, users activate the links by clicking on highlighted portions of their screen. This instantly displays the screen for the linked document.
IM	Information management.
Internet	International Communications Network.

Intranet	A limited access network that uses the Internet to link computers together. Access is limited to those individuals whose computers are on the Intranet network.
ITMRA	Information Technology Management Reform Act of 1996.
Java	A computer language for general purpose programming. Primarily used to write interactive Internet applications.
LAN	Local Area Network; a set of computers within a particular location (typically a building or part of a building) that have been physically connected together to enable shared communication and resources, such as printing, files, and dial-out capabilities.
LIPs	Laboratory Implementation Plans; detailed planning documents prepared by ORD Laboratories for all research to be conducted in-house or under a contract or cooperative agreement. Each LIP provides detailed information on the tasks to be conducted under a specific research project, the resources needed, and the research products that will be generated.
metadata	Data or information that describe a set of data or information. For example, scientific metadata describe how, when, and where the scientific data were collected.
model	Mathematical or physical representation of data or a system that accounts for all or some of its known properties.
NASA	The National Aeronautics and Space Administration.
NOAA	The National Oceanic and Atmospheric Administration.
object	A data or information product. For example, objects in ORD's ScienceNet may include data sets, databases, projects, analytical products, and documents.
OIRM	EPA's Office of Information Resources Management.
OMIS	ORD Management Information System; an ORD-wide integrated computerized management information system containing several modules for planning, financial management, project tracking, and human resources management.
QAPPs	Quality Assurance Project Plans; research planning documents, developed to ensure that data of the appropriate type and quality will be generated.
ScienceNet	The name ORD has selected for its Internet-based network of publicly available ORD science and engineering data and information. The ORD ScienceNet will be ORD's home page and front door on the Internet. It will contain printed documents, reports, and journal articles, as well as databases, data sets, models, and software applications. ORD hopes that the ORD ScienceNet will catalyze the creation of two larger scientific information networks: an EPA-wide ScienceNet and a broader U.S. ScienceNet developed collaboratively by EPA and other federal research agencies.
server	See "client/server."
STAR	Science To Achieve Results; ORD's external research grants program.
USGS	U.S. Geological Survey.
WAN	Wide Area Network; a group of Local Area Networks that have been connected together independently of the Internet to enable shared communication and resources, such as printing, files, and dial-out capabilities.
WWW	World Wide Web; a hypertext system for finding and accessing Internet resources.

Summary

This document, the *Information Management Component* of the Office of Research and Development's (ORD's) Strategic Plan, describes a consistent, organizationwide approach for efficiently planning for, collecting, documenting, manipulating, exchanging, archiving, and communicating ORD's research data and science information products.

As the science arm of the Agency, the results of ORD research provide the scientific foundation for environmental decision-making at EPA. With increasingly complex and multidisciplinary environmental research problems and programs, a new level of interaction and collaboration is required between different fields, specialties, scientists, and organizations that formerly may have operated independently. ORD's success as a federal-level R&D organization within this changing environment is defined by two factors:

- The scientific quality of our R&D.
- The degree to which our many EPA and external stakeholders and research partners have open access to and can use the information and data we generate.

ORD's IM Strategic Plan is structured around four themes that provide a robust framework for transforming ORD into a state-of-the-art information provider:

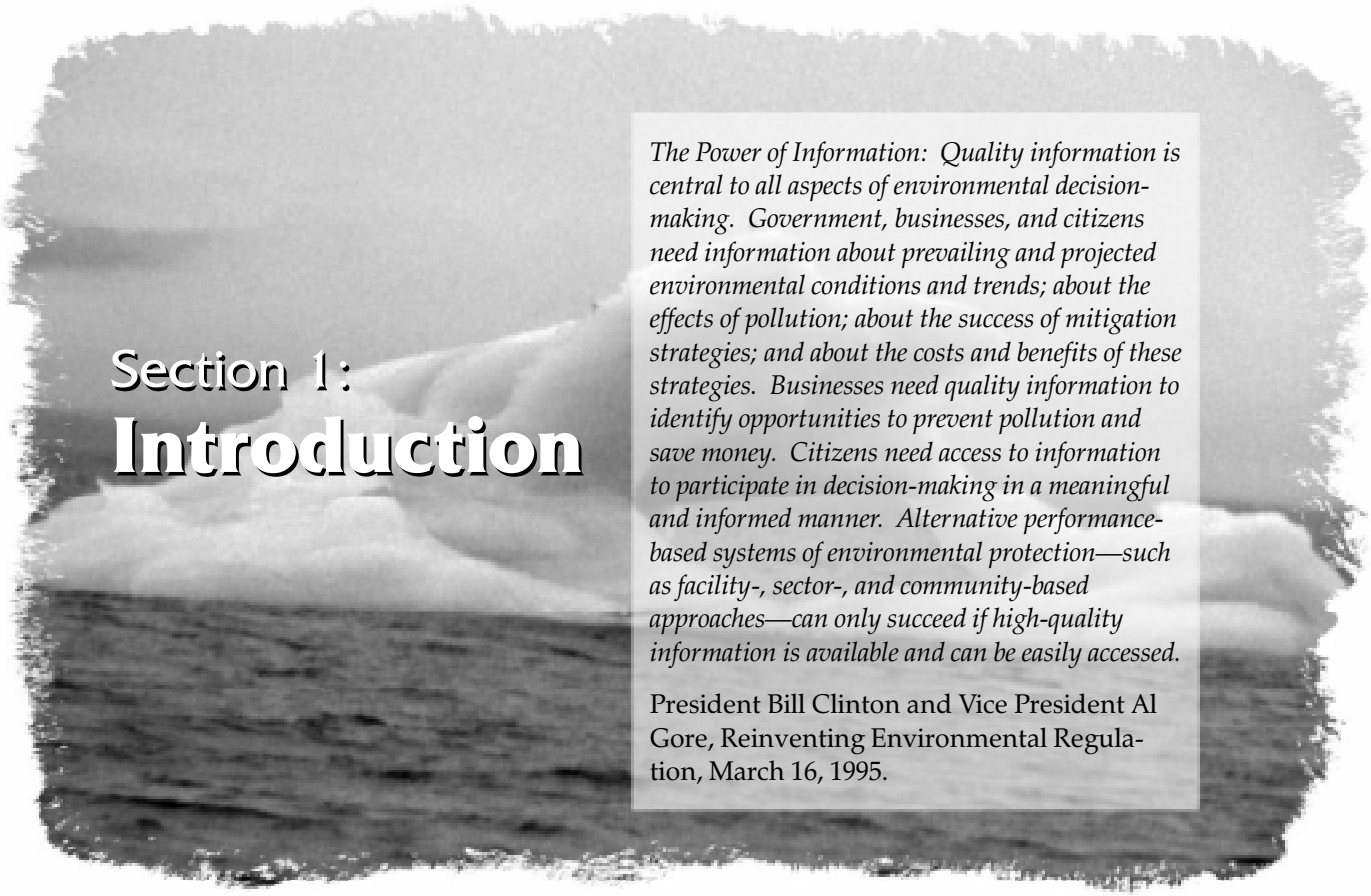
- **Planning** for information management.
- Making potential users **aware** that information exists.
- Making the information **accessible.**
- Making the information **usable.**

The strategic approach is comprehensive, offering a solution to ORD's wide range of information management needs, and addresses both paper and electronic information; interoperability and data management issues; policies and standards; data users and user needs; and electronic information technologies.

Central to this information management framework is the creation of an Internet-based **ORD ScienceNet,** the "front door" to ORD's scientific information. Through ScienceNet, a wide range of users—including researchers and data managers at other federal agencies, extramural scientists at all kinds of non-government laboratories, and the public at large—will be able to search for, access, and download data and information, as well as aggregate, manipulate, and analyze downloaded data sets using a variety of tools available via the World Wide Web. The ability to aggregate data may particularly aid environmental decision-makers, who often must synthesize large amounts of information during the decision-making process. For scientific data, users will be able to access metadata that describe the context and assumptions under which the data were collected. Also, users will be able to link from the ScienceNet website to other information or scientific databases (such as STORET and Envirofacts) via hypertext links.

The newly formed **ORD Science Information Management Coordination Board** will begin developing implementation plans for this strategy in August 1997. The Board will also be recommending an appropriate FY 1998 and 1999 budget to support high-priority ORD science information systems infrastructure.

Successful implementation of this strategy will move ORD from an organization that uses information technology for administration to one that uses it to further its scientific mission.



Section 1: Introduction

The Power of Information: Quality information is central to all aspects of environmental decision-making. Government, businesses, and citizens need information about prevailing and projected environmental conditions and trends; about the effects of pollution; about the success of mitigation strategies; and about the costs and benefits of these strategies. Businesses need quality information to identify opportunities to prevent pollution and save money. Citizens need access to information to participate in decision-making in a meaningful and informed manner. Alternative performance-based systems of environmental protection—such as facility-, sector-, and community-based approaches—can only succeed if high-quality information is available and can be easily accessed.

President Bill Clinton and Vice President Al Gore, Reinventing Environmental Regulation, March 16, 1995.

The Office of Research and Development (ORD) of the U.S. Environmental Protection Agency (EPA) is unique among scientific institutions in this country in combining research, analysis, and integration of scientific and engineering information across the full spectrum of health and ecological issues and across both risk assessment and risk management. ORD is responsible for leadership in science at EPA and for the bulk of EPA's R&D work. The results of ORD research provide the scientific foundation for environmental decision-making at EPA. They also are used by other government agencies, academia, nongovernment organizations, and the private sector for environmental science and management purposes at local, state, regional, national, and international levels.

ORD Success Factors

ORD's success as a federal-level R&D organization is defined by two factors:

- The scientific quality of our R&D.
- The degree to which our many EPA and external stakeholders have open access to and can use the information and data we generate.[1]

The recent explosion of tools and systems for information management and dissemination—relational databases, client/server computing, the Internet and World Wide Web, and compact disks, for example—offer extraordinary opportunities to enhance both the quality and application of ORD science. As never before, effective information management has the potential to:

- Increase ORD's R&D efficiency, quality, and accountability, sharpen the focus and value of our research, and catalyze rich collaborations with leading environmental researchers around the world.
- Vastly enlarge the constituency and potential applications for the data and information generated by ORD research. Recent technological advances make it possible for ORD to provide more usable information and data of documented quality, to more customers in more compelling formats, in a more timely manner, and at a lower cost than ever before.

[1]Throughout this strategy document, the term "ORD information" is used to signify the entire spectrum of ORD's science and administrative data and information.

ORD Information Management Strategy Group

In 1995, ORD instituted dramatic changes to transform itself into a world-class research institution to support risk-based decision-making. Specifically, we restructured our organization and instituted a new research planning process based on the widely utilized risk assessment/risk management paradigm. Our new risk-based planning process is described in our *Strategic Plan for the Office of Research and Development*, published for the first time in May 1996, and updated in April 1997.

To further enhance the quality and value of our work, we now have charted a course for ORD to take strategic advantage of the powerful information management opportunities enabled by recent technological developments. In February 1996, at the request of ORD's Deputy Assistant Administrator for Science, ORD created an Information Management Strategy Group consisting of representatives from all of ORD's Laboratories and Centers, the EPA Administrator's Office, and the EPA Office of Information Resources Management (see inside back cover). This group was charged with developing an innovative and coherent strategic plan for information management in ORD. This document, which is a formal addendum to ORD's Strategic Plan, sets forth that strategy.

Written for ORD staff and our EPA and external stakeholders, this plan provides a clear and practical blueprint for transforming ORD into a state-of-the-art information provider. This strategy is firmly grounded on a set of fundamental economic, technological, and strategic principles to ensure that ORD's approach to information management will be realistic, successful, and take maximum advantage of available resources and opportunities.

The plan begins by defining a vision, mission, and goals for ORD information management (Section 2) and examining the needs of ORD staff and stakeholders (Section 3). Based on these needs, the plan sets forth a four-part strategy for managing information and defines a number of specific actions ORD will take to implement the strategy (Section 4).

Finally, this document will guide detailed implementation planning, which will begin in late summer, 1997. A newly formed ORD Science Information Management Coordination Board, composed of senior information officers from each ORD Laboratory and Center and chaired by a member of ORD's Executive Council, will initially focus on:

- Development of plans for implementation of this strategy; and
- Recommendations for appropriate allocation of resources to be set aside in FY 1998 and 1999 to support high-priority ORD science information systems infrastructure.

Strategic Principles for ORD Information Management

- Build on and coordinate ORD's existing information management capabilities.
- Be ORD-wide and flexible (rather than monolithic or centralized).
- Minimize the cost and disruption to ORD's current operations.
- Be integrated with Agency-wide information management planning.
- Take advantage of rapidly evolving information management tools.
- Leverage resources from other agencies.

External Review

On March 20, 1997, an outside team of experienced information systems managers (drawn from NASA, NOAA, USGS, CENR, and EPA's OIRM) conducted a one-day review of this strategy. The review committee stated that they were very impressed with the quality and commitment within ORD to this effort, and made a number of insightful comments and suggestions for refining our information management planning and implementation. Their recommendations have been incorporated into this strategy.

Coordination of This Strategic Plan[2]

Both ORD and the Agency as a whole already have made significant investments in information management systems. This strategy builds on those efforts by proposing an approach based on organizing and enhancing existing ORD and EPA systems and resources.

The proposed approach is also comprehensive, serving as a unifying plan of action for managing all levels and types of ORD information—from the

[2]It will be important for the new ORD Science Information Management Coordination Board to continue these efforts, by coordinating implementation planning with relevant activities and organizations outside of ORD (Section 3).

scientific data and information resulting from ORD's in-house or extramural research (e.g., raw data collected at field sites, health or ecological risk assessments, aggregated data sets, research plans) to the administrative information needed to manage ORD's research (e.g., resource data, grant award information, Laboratory Implementation Plans).

In addition, this strategy has been coordinated and is consistent with relevant organizations and related activities, such as the Agency Information Resources Management Strategic Plan and the new Office of Planning, Analysis, and Accountability. It is also responsive to the Government Performance and Results Act (GPRA) and the Information Technology Management Reform Act, which was signed into law in February 1996 and requires a fundamental change in the way government agencies perform information management.

The Information Technology Management Reform Act

The Information Technology Management Reform Act (ITMRA) of 1996 establishes a set of information technology acquisition and management requirements for government agencies.

These requirements, summarized in Appendix A, are designed to maximize the value of government investments in information technology while minimizing the risk.

One ITMRA requirement is that EPA establish a Chief Information Officer for the Agency (currently the Acting Assistant Administrator for Administration and Resources Management), responsible for ensuring Agency-wide compliance with ITMRA. Concomitantly, ORD must ensure that its IM activities and systems conform with the ITMRA requirements. Essentially, this means approaching each potential information technology investment as a business case and apply rigorous cost-benefit analyses to it.

In response to ITMRA, EPA already has:

- Initiated an Agency-wide Strategic Information Technology Architecture Plan process.
- Established a Capital Planning and Investment Control Strategic Project Committee to define EPA's process for capital planning and investment control, determine best practices, and implement a new method of evaluating information technology investments, as required by ITMRA.

To keep pace with and help shape these developments, ORD must proceed without delay to participate in these efforts.

Next Steps

The time to act is now. Other EPA programs and offices are currently exploring new information management concepts, such as electronic permitting, reporting and pesticide registration, and "virtual government" pilot projects are being discussed with other federal agencies to develop better interagency connections. At this critical juncture, ORD has the opportunity to lead the Agency in the development of a coordinated information network for science—an "ORD ScienceNet" (described in Section 4)—at a time when the need for defensible and credible environmental research information has never been greater. Given that so much is now possible in electronic networking and publishing, this is an exciting time for ORD to make dramatic progress by "leapfrogging" over obsolete technologies and approaches for information management and into cutting edge systems that hold dramatic promise for enhancing both ORD's and the Agency's stature as a world-class environmental institution.

Section 2: ORD's Vision, Mission, and Goals for Information Management

Many of EPA's scientific data sets are either incomplete, obsolete, or missing altogether, a problem that extends across all media areas. These problems have made it difficult for the Agency to conduct scientifically based risk assessments and to measure the results of environmental programs. The Agency's problems in obtaining quality data are exacerbated by difficulties in managing the data that are available. In particular, EPA's data management systems have been designed to track or manage information about environmental conditions and results for each environmental medium, thus making scientific assessments of risk across media cumbersome and costly.

Peter F. Guerrero, GAO, May 12, 1995, testimony before the Senate Subcommittee on VA, HUD, and Independent Agencies.

Introduction

The evolution of environmental science in general and ORD's program in particular, is distinguished by several trends that have important implications for information management. First, research problems and programs are becoming increasingly multidisciplinary, requiring collaborations and interaction between different fields and specialties that formerly may have progressed independently. Second, environmental science in the U.S. is benefiting from growing coordination and partnering among multiple federal agencies, each of which is bringing its own resources and research communities to contribute to more integrated federal efforts. Third, ORD's development of a broader extramural component to its research efforts, e.g., through the Science To Achieve Results (STAR) program of research grants, means that the community of scientists who are playing key roles in many research programs now are located at many different institutions in addition to ORD's own laboratories. And finally, there is a growing interest on the part of researchers who are not directly involved in the research programs for which data are initially produced, and on the part of the general public, to have access to EPA data sets. The consequence of all these trends is that ORD's approach to IM must place a priority on making data openly available to a wide range of users, including researchers and data managers at other federal agencies, extramural scientists at all kinds of non-government laboratories, and the public at large.

In its 1996 Strategic Plan, ORD set forth a new organizational vision, mission, and goals to provide the framework for risk-based research planning.

Similarly, ORD has developed a vision, mission, and goals for information management. This framework is designed to ensure that ORD manages information in the way that will best support risk-based research planning and the new trends in environmental decision-making. As described below, the IM framework directly parallels and supports ORD's broader organizational vision and mission.

Vision Statement

ORD's vision for information management is based on the key role that ORD science plays within EPA, and within the broader context of our nation's environmental research agenda. Building upon

ORD's organizational vision statement—*"ORD will provide the scientific foundation to support EPA's mission"*—ORD will likewise support the Agency's mission by conforming to its vision for information management:

ORD's Information Management Vision

ORD will provide timely and reliable scientific data to support EPA's mission, and will exchange environmental information with the public and other stakeholders.

The ORD information management vision statement provides a standard against which the future ORD information management environment can be judged. By living up to this standard, ORD will support the Agency's mission and will also serve as a public resource for reliable scientific, engineering, and risk assessment/risk management information. At the same time, ORD will increase its capacity to exchange environmental information with and integrate the work of ORD's scientific partners, including EPA's Program and Regional Offices, academia, the private sector, and other government agencies.

Mission Statement

As EPA's science arm, ORD has a clear role to identify and provide defensible and credible data and information. Such information is critical to achieving the Agency's mission of protecting human health and the environment. Effective environmental decisions depend on scientific and engineering data that are accurate, reliable, and adequate for their intended use (see "QA Program Interface"). And defensible environmental decisions are backed up by documentation of both the science data and the risk assessment procedures and analytical methods that were used to develop them. ORD plays a leadership role in developing these tools and information and in providing, coordinating, and exchanging information with decision-makers inside and outside the Agency. ORD also supports Agency-wide internal information and data needs for strategic planning, budgeting, and accountability.

The ORD IM mission statement parallels the four-component structure of ORD's overall mission statement, set forth in the Strategic Plan for ORD:

ORD's mission is to perform research and development—

- ORD's IM mission is to make internal and external stakeholders aware of and able to access and use the data and information generated by ORD's research and development.

ORD's mission is to provide technical support—

- ORD's IM mission is to provide research information and supporting electronic communications to customers in a manner that meets world-class standards for quality of content and delivery.

ORD's mission is to integrate scientific research—

- ORD's IM mission is to foster interactive communication, collaboration, and information sharing with scientific partners.

ORD's mission is to provide scientific leadership—

- ORD's IM mission is to provide leadership in the Agency's use of information technology for science.

The parallel structure clearly indicates how ORD will use information management to support its organizational mission and reflects ORD's critical role as an information provider.

Goals

This strategy is structured around four themes that provide a robust framework for effective ORD information management:

- **Planning**—IM planning must be an *integral component of ORD research planning* to ensure that the information generated by all of ORD's research projects can be made available to potential users.
- **Awareness**—Once information is available, stakeholders need to be *aware* that it exists.
- **Accessibility**—Then, they need *access* to the information.
- **Usability**—Finally, the information must be *usable.*

These four themes form the basis for a set of broad, customer-oriented IM goals (Table 2-1) to fulfill ORD's IM mission. Section 4 describes a strategy to achieve these goals.

Table 2-1. ORD's Long-Term Goals for Information Management

Planning	Awareness	Access	Usability
ORD will integrate IM planning into its research planning process to ensure that all of ORD's research information can be made available to potential users in a timely, effective, and efficient manner.	ORD will provide the awareness tools and services needed to make its internal and external stakeholders aware of ORD's information.	ORD will provide the communication paths and equipment that allow internal and external stakeholders access to ORD's information.	ORD will provide the policies and standards, training, user tools, and planning needed to make ORD's information usable to internal and external stakeholders.

Section 3: ORD Information Management Needs, Opportunities, and Benefits

Scientists can now perform environmental research that increases our understanding of the Earth system at all spatial scales, enhances resource management and environmental decision-making, and improves our capabilities for predicting changes in the environment. Over the past decade, in particular, the observational, computational, and communications technologies have enabled the scientific community to understand a broad range of interdisciplinary environmental research and assessment programs.

The National Research Council, Finding the Forest in the Trees.

Needs

ORD currently employs almost 1,900 scientists, engineers, and professional and support staff. These individuals perform or support research in human and ecological health, methods and model development, and risk management. ORD staff also provide technical support to EPA's Program and Regional Office staff as well as external stakeholders.

These activities generate four categories of information, each used primarily by different audiences (Table 3-1):

- Scientific information (risk assessments, research products, the Integrated Risk Information System, and so on)—primarily used by **ORD stakeholders.**
- Integrated scientific, management, and administrative information (e.g., planning, budgeting, and accountability information)—primarily used for making strategic, policy-level decisions by **ORD senior executives.**
- Science management information (i.e., information related to managing and administering ORD research including planning and resource information)—primarily used on a more tactical level by **ORD project managers and administrative personnel.**
- Science data (e.g., monitoring data, analytical data, epidemiological data, and so on)—primarily used in the development of research products by **ORD scientists and engineers.**

As described below, each of these user groups has specific information management needs that, when fulfilled by the strategy presented in Section 4, will greatly enhance their ability to access and utilize information that directly or indirectly contributes to ORD's ability to be a world-class environmental research institution.

ORD User Council

The degree to which our many EPA and external stakeholders use the information and data we generate is listed in Section 1 as one of two ORD success factors. It is critical, therefore, for ORD to

Table 3-1. The Need for ORD Information*

Who needs ORD information?	ORD Stakeholders	ORD Sr. Executives	ORD Project Managers & Administrative Personnel	ORD Scientists & Engineers
What information is needed?	Science Information Products	Integrated Scientific, Management & Administrative Information	Science Management Information	Science Data
Examples:	• Risk assessments • Research publications • Integrated Risk Information System (IRIS)	• Planning information • Budgeting information • Accounting information	• Planning information • Resource information • Laboratory Implementation Plans • Grant award information	• Human health & ecological effects data • Human health & ecological exposure data • Risk management data • Methods & models data • Monitoring & analytical data

* Examples are given only to illustrate that there are several broad categories of ORD information that, for the most part, are needed by different user groups. Some overlap between categories does occur.

have external user communities actively involved with our information management program. These stakeholders must be involved from the outset in ORD's planning, so that they buy into the process and their needs for information are met.

Because of their key role, ORD will establish a "User Council" to ensure that we adequately coordinate with and provide for the information needs of our external user community. The appropriate membership of the User Council and other specifics will be addressed by the ORD Science Information Management Coordination Board during implementation planning.

ORD Stakeholders

EPA Program, PAA, and Regional Offices
ORD is the primary provider of scientific and technical information, as well as technical assistance, to EPA's Program and Regional Offices. ORD staff need ready access to the information and data necessary to meet the information and assistance needs of the Program and Regional Offices. The needs and activities of the Agency's Planning, Analysis, and Accountability Office also will need to be met. In addition, as appropriate, the Program and Regional Offices should have direct access to relevant information and data generated by ORD.

Federal Entities (Agencies, Tribes, Institutions, and Committees)
Federal entities, including agencies, institutions, tribes, and committees, are key ORD partners and stakeholders in several ways:

- Federal research institutions—such as the National Academy of Sciences, the National Science Foundation, the National Aeronautics and Space Administration, and the U.S. Department of Energy—are important research partners of ORD. Often, these agencies help establish standards for best uses of information technology for research purposes.
- ORD represents EPA on federal-level committees, such as those of the National Science and Technology Council, which coordinates the research activities of the federal government.
- ORD also participates in major interagency research initiatives—such as the work of the Federal Geographic Data Committee to build the

National Spatial Data Infrastructure, and the National Biological Services effort to build the National Biological Information Infrastructure.

- Environmental research information forms an important pillar of the National Information Infrastructure and the Global Information Infrastructure.
- Finally, ORD and tribes have collaborated on research.

Current trends in government policy and technology are driving ORD toward greater networking with its federal research partners and customers. This networking is essential to exchange information and coordinate activities so that ORD and other federal institutions can most effectively leverage their increasingly scarce resources. ORD needs to learn from the experiences and efforts of the other research agencies and make its information management program synergistic with theirs.

States and Communities

Increasingly, the Agency is moving toward facilitating regional, state, and local action to identify and solve environmental problems where appropriate and feasible. For example:

- Facilitating and supporting stakeholder partnerships for environmental improvement is one of the key strategic directions cited in EPA's Five-Year Strategic Plan, published in 1994.
- The National Advisory Council for Environmental Policy and Technology (NACEPT), a federal advisory committee, has been providing recommendations to the EPA Administrator on how the Agency can most effectively catalyze community-based environmental protection efforts. NACEPT's 1996 Community-Based Environmental Protection Committee has identified information as one of the primary ways that EPA can catalyze community-based environmental protection (CBEP) efforts.

As ORD works to meet information management needs, it should consider what types of ORD information may be valuable for state and community-led environmental protection efforts and find ways to make this information readily accessible to these stakeholders. Remote sensing and GIS data may be of particular value to these stakeholders.

ORD Senior Executives

ORD's executive leadership provides strategic, policy-level decision-making for ORD. The senior executives define ORD's vision, goals, and objectives, determine how best to ensure the success of ORD's efforts, and define priorities for the content and direction of ORD's research program. They also identify critical gaps in ORD's management or science agenda and take steps to address those gaps.

ORD senior executives need a variety of science, management, and administrative information to support effective decision-making. ORD has a finite set of resources that can be allocated in different ways to support activities critical to EPA's mission. The management challenge for senior executives is to focus ORD research activities on those areas where the greatest benefit can be achieved most efficiently and cost-effectively. With the right kinds of information, this type of decision-making is facilitated and defensible.

To move up to the next generation of strategic research planning and directing, ***ORD senior executives need not only ready access to administrative information, but will need the ability to integrate it with ORD's science information.*** Examples of improved research planning facilitated by using integrated administrative and science information as a strategic resource include:

- Allocation of resources to specific geographic initiatives.
- Knowing what data are available and what projects are ongoing in a specific geographic region may allow improved decisions on where research can be most efficiently conducted.
- Choosing an area already rich in relevant, well-documented data may reduce the resource needs for data collection efforts and significantly shorten the time needed to produce project results.
- Knowledge of ongoing projects within EPA and other organizations.
- Access to information describing projects will significantly reduce the potential for redundant research efforts on the part of ORD. This type of information can be extracted from administrative data describing internal projects, grants, contracts, and interagency and cooperative agreements. Similar information for external entities can be found at their Internet sites.
- Knowledge of existing capability and capacity of EPA and external organizations.
- Access to information about expertise in EPA and other organizations can lead to the development

of strategic partnerships that will allow ORD to more efficiently accomplish its mission.

- Ability to access and evaluate research results in order to direct further activity.
- Access to information on research results will allow ORD to be more effective and timely in responding to Congressional and other inquiries.

ORD Program Managers and Administrative Personnel

As described below, ORD program managers and administrative personnel need tools and services that allow them to effectively:

- Plan and manage ORD research programs.
- Manage the financial and human resources that support the research programs.

Program Planning and Management Needs
To effectively plan and manage ORD research programs, ORD program managers and administrative personnel need:

- Information about ongoing research.
- Information about the needs and operation of the Agency's Planning, Analysis, and Accountability function.
- Reliable and easily obtainable resource information about intramural and extramural research projects.
- Information about publications and other end products of ORD research.
- In some cases, desktop access to scientific and administrative data on the Internet and other sources for use in risk assessment and risk management decision-making.

For example, when proposing or designing an ecological research project, an ORD program manager could access the RaDiUs (Research and Development in the U.S.) dedicated server to determine what research U.S. government agencies are conducting on that particular ecological issue. This information will allow the program manager to determine what funds have been allocated to this area of research and will facilitate networking with other researchers. The ready availability of this type of information will be invaluable for reducing overlap and duplication of effort and will enable ORD laboratories to target their research based on a comprehensive understanding of other similar research efforts.

Resource Management Needs
To effectively manage the financial and human resources that support ORD research programs, ORD program managers and administrative personnel need to:

- Develop effective Laboratory Implementation Plans.
- Link task-based tracking of operating budgets and products.
- Maintain an audit trail of how resources were reallocated.
- Plan and track travel and miscellaneous expenditures.
- Support the research planning process from planning, through budget execution, to results tracking.
- Provide complete records of intramural and extramural research plans, activities, and accomplishments.
- Help ORD keep "in sync" with the Agency's Integrated Financial Management System and the needs and activities of the Planning, Analysis, and Accountability Office.

In 1996, ORD launched the ORD Management Information System (OMIS), which has been designed to meet these needs. OMIS is an ORD-wide integrated computerized system containing several modules for planning, financial management, project tracking, and human resources management. The component modules are:

- The Integrated Resources Management System.
- The in-house accounting module.
- The human resources module.
- The Laboratory Implementation Plan module and other tracking system modules under development.

By the end of FY97, OMIS will hold complete and organizationally consistent information on ORD research projects and tasks; commitments, obligations, and expenditures; grants and contracts; and training plans. OMIS is expected to meet ORD's administrative information management needs and is designed to be consistent with all the Agency's budgeting and accounting requirements.

Concurrently, the Agency's administrative information management systems are being reformed. The

new or updated systems will provide useful tools for meeting ORD resource management needs. For example, under the ARI, or Administrative Reduction Initiative, EPA is adding a variety of new systems capable of electronic routing. These systems will allow ORD managers to electronically receive, review, approve, and act on repetitive processes—such as leave slips, executive correspondence, purchase requests, travel authorizations and travel vouchers, and time and attendance information.

Finally, ORD is piloting "groupware" computer software that will allow ORD program managers and administrative personnel to participate in the many initiatives that will become the standard for EPA administrative processes within the next 18 months to 2 years.

ORD Scientists and Engineers

To play a leadership role in environmental R&D, ORD scientists and engineers need information management capabilities that allow them to:

- Plan and discuss projects.
- Locate and understand data, information, and documents.
- Access data, information, and documents.
- Collect, store, manipulate, and analyze data and information.
- Publish research data and results.

The need for these capabilities is discussed below.

Need To Plan and Discuss Projects
Increasingly, cross-disciplinary teamwork is needed to conceive and conduct research that addresses complex and emerging environmental concerns, which often are multimedia in nature. To work effectively in teams, researchers need the capability to develop plans using an iterative process that incorporates input from multiple sources and encompasses both the strategic and tactical levels. There are three basic stages of IM needs during planning:

- ORD researchers located at different geographic locations need to be able to communicate and interact electronically when developing project concepts.
- Once project concepts are formulated, ORD researchers in different locations need an efficient mechanism to develop, discuss, and review Laboratory Implementation Plans containing milestones, deliverables, and task descriptions.
- Finally, when the research project is underway, ORD scientists and engineers need to consolidate updates from multiple sources to assess impacts and progress toward achieving project objectives on an ongoing basis.

ORD's local and wide area information management environment must support and facilitate planning activities at all three stages.

Need To Locate and Understand Data, Information, and Documents
ORD scientists and engineers need to work with their own data and with secondary data (i.e., data generated by others). To effectively use secondary data, ORD users must be able to locate the data and readily determine whether it is appropriate for their needs.

Increasingly, government agencies, nongovernment organizations, and academia are providing data and information on the Internet that is well-documented and potentially useful to ORD staff. Emerging metadata (information that describes data) standards like those adopted by the Government Information Locator Service (GILS) and the Federal Geographic Data Committee standard for spatial data documentation help ensure that these data and information are appropriately documented and can be readily located.

To stay abreast of the latest scientific developments, ORD scientists and engineers need the capability to:

- Search the Internet for relevant data.
- Perform more traditional literature searches via commercial or government services (such as those supplied by the National Institutes of Health and the U.S. Department of Agriculture).

Need To Access Data, Information, and Documents
Once ORD users have located appropriate data, they must be able to access the data, as well as metadata and documentation, using an appropriate tool set. Examples of access scenarios include:

- Selecting a particular geographic coverage on an Internet server, using an Internet client that supports file transfer protocol to download the selected coverage, and then displaying it using a geographic information system (GIS).
- Using a database browser tool, statistical package, or GIS to examine the content of data in a database management system.

- Using the wide variety of valuable scientific data, information, and documents available on CD-ROM.

Need To Collect, Store, Manipulate, and Analyze Data and Information

Much of ORD's data is originally generated by field and laboratory work that is either conducted or sponsored by ORD. These data need to be managed from their origin in the field or laboratory through the analytical process. To do this effectively, ORD users need access to a wide variety of computer tools, including laboratory information systems, field data recording equipment, and analytical tools. General categories of tools include modeling, scientific visualization, geographic information systems, statistical, spreadsheets, Internet and Intranet servers and browsers, and scientific database management tools.

Need To Publish Research Data and Results

ORD's success at meeting its mission will be largely judged on what it publishes. Forms of "publication" include journal articles, databases, data sets, metadata that describe the databases and data sets, models and software applications developed by ORD, and documents. ORD's Science To Achieve Results (STAR) program is currently generating a large body of research data and scientific results that will become part of ORD's information collection in the future.

Electronic publishing is important for internal and external review of draft materials and public access to final materials. To publish electronically for internal purposes, ORD staff need a robust capability that allows them to work in groups to develop, review, and track documents internally. To publish electronically for external purposes, ORD personnel must be able to work on the Internet.

Opportunities and Benefits

Until a few years ago, limitations of information technologies made it difficult, if not often impossible, to set up comprehensive, integrated, and effective IM systems across large organizations such as ORD. During the 1990s, however, information technology and management approaches have made revolutionary progress:

- Key technologies include the Internet and its associated software tools (especially the World Wide Web and Java).
- Emerging management approaches include the concept of creating a Chief Information Officer to oversee information management, as well as a number of other strategic IM approaches that government agencies are now required to implement under the Information Technology Management Reform Act of 1996 (see Appendix A).

These new technologies and approaches provide ORD with important opportunities to achieve the six long-term goals set forth in ORD's 1996 Strategic Plan (Table 3-2) and to meet the information management needs described in this section. In particular, they enable ORD to:

- Integrate its many discrete databases.
- Integrate its information activities with the rest of EPA.
- More readily communicate with and provide information to external partners and stakeholders.

These new opportunities have the potential to bring many benefits both to ORD staff members and to ORD as a whole. For example:

- ORD researchers will be better able to collaborate with research partners, conduct multidisciplinary research, and use GIS and other emerging software tools to readily visualize and make sense of large quantities of data. Sharing of data sets with others will help to correct errors and to clear up misunderstandings in the metadata.
- ORD researchers will have increased awareness of ORD data, and better quality data sets and metadata will lead to better research.
- Better long-term archiving and documentation will protect data sets for future use by ORD and other researchers.
- ORD scientists and engineers will gain substantially increased visibility for and recognition of their work.
- Improved information management has substantial potential to enhance ORD's stature, improve decision-making and accountability, and enhance efficiency and employee satisfaction.

In short, enhanced information management is imperative if ORD is to serve as a world-class provider of environmental data and information in the next century. Section 4 of this document sets forth an ORD IM strategy to effectively meet the many needs of ORD staff and stakeholders described above.

Table 3-2. ORD's Long-Term Goals

- To develop scientifically sound approaches to assessing and characterizing risks to human health and the environment.
- To integrate human health and ecological assessment methods into a comprehensive multimedia assessment methodology.
- To provide common sense and cost-effective approaches for preventing and managing risks.
- To provide credible, state-of-the-art risk assessments, methods, models, and guidance.
- To exchange reliable scientific, engineering, and risk assessment/risk management information among private and public stakeholders.
- To provide leadership and encourage others to participate in identifying emerging environmental issues, characterizing the risks associated with these issues, and developing ways of preventing or reducing these risks.

Benefits to ORD of Enhanced Information Management

- **Improved research**. The potential benefits to ORD research of enhanced IM are exponential:

  - ORD researchers will be able to stay abreast of important scientific developments in a timely and efficient manner.
  - ORD research will be enriched by new opportunities for data exchange and collaboration with internal and external researchers.
  - Enhanced sharing of data sets will stress the importance of adequate metadata and ultimately lead to metadata improvements.
  - Better long-term archiving and documentation will protect data sets for future use.
  - Use of ORD data sets will increase substantially due to improved accessibility, preservation, and metadata quality.

- **Increased stakeholder appreciation of ORD's work**. Using new IM technologies, ORD can now make its products available to stakeholders in the U.S. and around the world more efficiently and effectively than ever before. Both ORD and stakeholders benefit:

  - Many more stakeholders will be able to readily identify, access, and use ORD tools and information for environmental research, decision-making, and improvement.
  - Taxpayers get a better return on their investment through enhanced application of ORD's work.
  - External accessibility to ORD products brings positive recognition and increased support for ORD's work. NASA gained tremendous public visibility and appreciation for its work by placing Hubble telescope images of Jupiter on the Internet and by providing a forum for public interaction on the event. ORD could use the Internet for similar purposes.

- **Greater efficiency and cost-effectiveness**. Improved access to key information will enhance decision-making at all levels of ORD. It also will facilitate effective planning that provides maximum leverage of ORD's resources. This will help ORD conserve resources and become more cost-effective.

- **Enhanced accountability**. Enhanced management of ORD scientific, management, and financial information will provide an essential foundation for transparency and accountability to stakeholders. Also, better documentation of ORD's data holdings will improve the defensibility of ORD and EPA decisions.

- **Enhanced employee understanding and satisfactions**. By publishing the ORD employee handbook and other key resources on-line, ORD can better guide employees through its processes and organization. Also enhanced IM will provide ORD employees with tools and information that offer important opportunities for improving performance and efficiency.

Section 4: Strategy for Managing ORD's Information

EPA must have the ability to collect, process, and analyze the information needed to ensure that it is managing for and achieving real environmental results.

EPA's Five-Year Strategic Plan: The New Generation of Environmental Protection, U.S. EPA, July 1994.

Intoduction

This section describes ORD's four-part strategy for managing its information that is based on several fundamental tenets of successful information management (Figure 4-1):

- **Planning** for information management.
- Making potential users **aware** that information exists.
- Making the information **accessible** to users.
- Making the information **usable.**

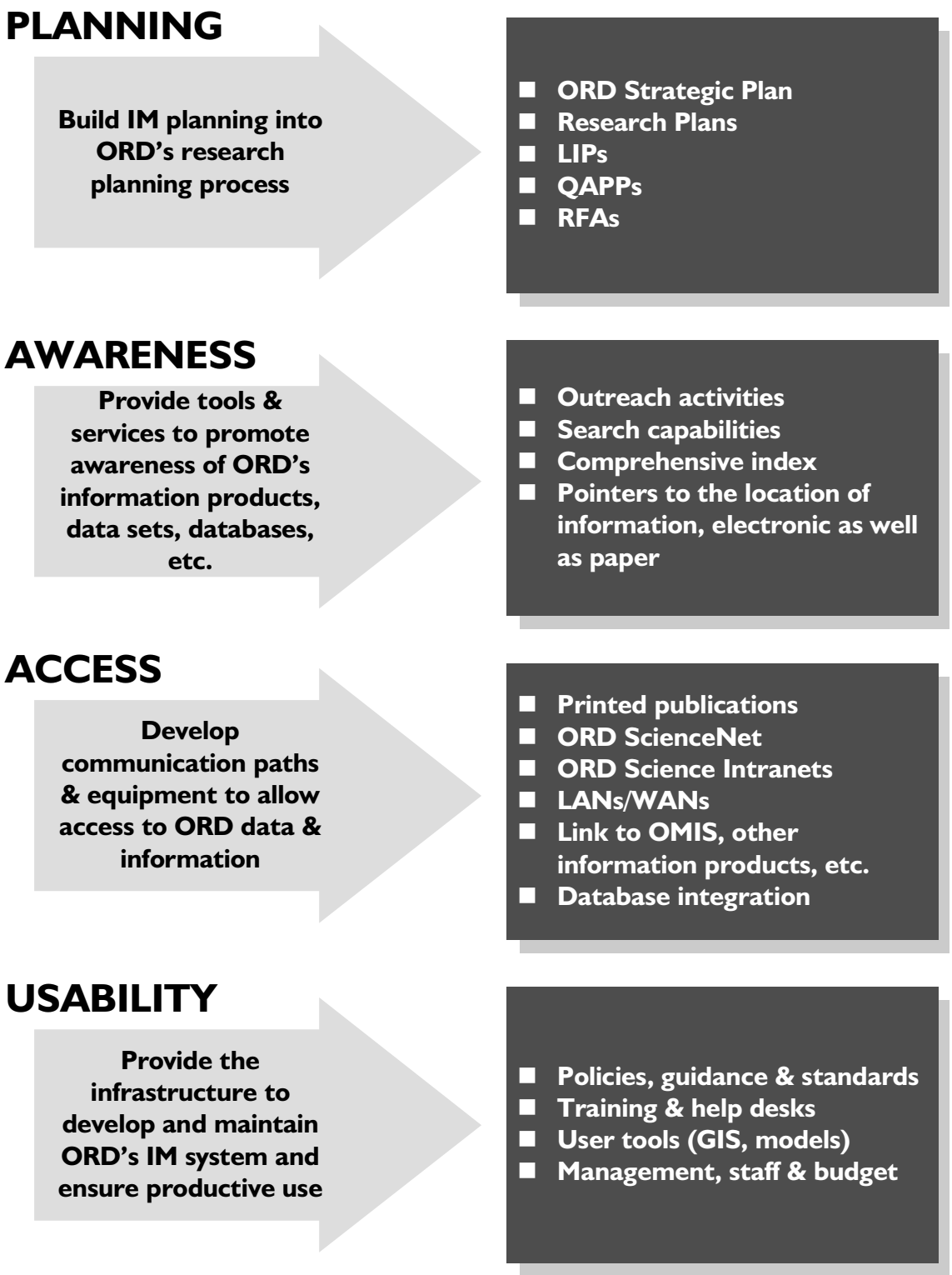
Described in this section is an approach for information management that:

- Will meet ORD's information management needs (described in Section 3) using a comprehensive approach that addresses: both paper and electronic information; research planning; science data management; policies and standards; user needs and support services; and electronic information technologies.
- Takes advantage of the unprecedented information technology opportunities now available.
- Provides a consistent, organizationwide approach to enable ORD to efficiently plan for, collect, document, archive, manipulate, and distribute scientific and management information—including the information generated by both intramural and extramural research.
- Builds on and integrates ORD's existing information capabilities—including tools, services, communication paths, communication equipment, and standards.

Strategic vs. Implementation Plans

ORD's plan for information management is presented in this document at a strategic level, focusing on the *direction* ORD intends to take, not *specifically how* it will be accomplished. Specifics will be left, for the most part, for implementation plans that will be developed and coordinated across the organization by the ORD Science Information Management Coordination Board (Section 1). These implementation plans will spell out exactly how ORD will execute this strategy.

Figure 4-1. Strategic Approach.*



* ORD's information management strategy is not an "Internet plan"—it is a comprehensive approach that addresses "paper-based" and electronically stored information; research planning processes; science data management; administrative information systems; policies and standards; and user needs and support services—as well as information technologies.

However, a few specific implementation items are included in this Strategic Plan, either to provide concrete examples so that the reader has a better understanding of the issue being discussed, or because noteworthy progress has already been made in a certain area.

Planning: Strategy Component 1

As mentioned in Section 1, ORD currently has in place a strategic planning process that determines the highest priorities for intramural and extramural research based on risk assessment and risk management principles. This process was finalized in 1996 and published in the *Strategic Plan for the Office of Research and Development*. The 1997 edition of the Strategic Plan was issued in the spring of this year, and now includes information management considerations.

The vast majority of the data and information that ORD will make available under this IM strategy stems directly from the process and products of ORD research. Therefore, as described below, planning for information management must become an integral part of ORD's research planning process.

ORD's Strategic Planning Process

ORD's risk-based planning process is described in Appendix B. The process involves soliciting stakeholder input, identifying potential research topics, evaluating these topics to select the most appropriate ones for ORD's research agenda, and then defining specific research needs for the selected topics.

Once ORD has identified specific research needs, ORD staff develop a series of planning documents, including Research Plans, Laboratory Implementation Plans, Quality Assurance Project Plans, and Requests for Applications (see the section after next for more detail). Research then is conducted based on these plans. The research products are delivered to stakeholders and provide input into the next strategic planning cycle.

Planning for Information Management

Three considerations are paramount if ORD is to successfully integrate IM planning into the research planning process:

- **Begin IM Planning Upfront—**Information management planning should commence *as soon as ORD has identified its specific research needs.* This is the optimal point for ORD to begin planning for the data and information that will be generated by the specific research projects.
- **Include a Budget for IM—**ORD management and budget decisions must be made *considering the entire research project*, from data collection through long-term archiving of data sets.
- **Consider All Research Types—**Information management planning must, to differing degrees, encompass *all* ORD research projects, including in-house research and extramural research that ORD funds with contracts, cooperative agreements, and grants.

Information included in Appendix B describes the portions of the overall ORD research planning process that should involve planning for information management.

Examples of IM Planning Questions to Consider at All Stages of ORD Research Planning

- What is the purpose or use for which the data were collected?
- What data quality indicators were met supporting the data quality objectives?
- How much data will the program/project generate?
- Who and where are the data users?
- What are the possible long-term archival uses?
- Is there an archive ready to accept and handle the data?
- Who will set standards, delivery requirements, milestones, etc.?
- Who will deliver, document, manage, and distribute the data?
- Are data sets part of deliverables or products for ORD partners or clients?
- What are the requirements for added-value processing? What computing, data management, distribution capabilities are needed? Where do these capabilities reside?
- Who will budget for IM and who will allocate resources among participants?
- How will evolutions in IM technology be accommodated over the life of the project (this question is particularly relevant for long-term projects that extend over many years)?
- Who will set the policy on data access?

When planning for information management, ORD should be guided by the awareness, access, and usability principles (described below) of this IM strategy. In other words, ORD research planners should consider not only what information the research will generate, but also how best to:

- Make potential information users (including clients and stakeholders) aware that the information exists.
- Describe, organize, and display the information so that it will have maximum accessibility and utility to users.

This involves considering who will use the data and how, defining a schedule and resources for IM activities, finding the appropriate "home" for the information, and establishing standards and access policies. The text box on page 19 provides examples of the types of questions ORD managers should address when planning research.

Integrating IM Planning Into ORD Research Planning Documents

During research planning, ORD develops several types of documents that set forth the scientific questions to be addressed and specify schedules, milestones, resources, quality assurance, peer review requirements, and so on. These include:

- *ORD Research Plans*, which provide a broad, overall context and strategic direction for ORD research projects/programs.
- *Laboratory/Center Research Plans*, which are prepared by ORD's National Laboratories and Centers as planning tools to identify, prioritize, and justify their overall research programs.
- *Division and Cross-Division Research Plans,* which define, prioritize, and justify the research program to be conducted by a single division within an ORD Laboratory or Center or by multiple divisions collaborating within or across ORD Laboratories and Centers.
- *Laboratory Implementation Plans*, which provide detailed information on discrete tasks for use by ORD in planning and managing research conducted in-house or under a contract or cooperative agreement.
- *Quality Management and Quality Assurance Project Plans (QMPs & QAPPs),* which include sections on assuring the appropriate type and quality of scientific and engineering data.
- *Requests for Applications (RFAs)*, which invite research grant applications from the external scientific community in areas of special interest to EPA's mission.

Figure 4-2 shows the relationship of these ORD planning documents.

To ensure coordinated information management within ORD, each of these documents should explicitly specify how the data and information to be generated by the research will be managed. Upfront information planning will help ensure that the results of ORD research are captured and disseminated efficiently, rapidly, and as broadly as is appropriate in a way that will be most useful to science data users. Appendix B describes specific recommendations for each type of ORD research planning document.

QA Program Interface

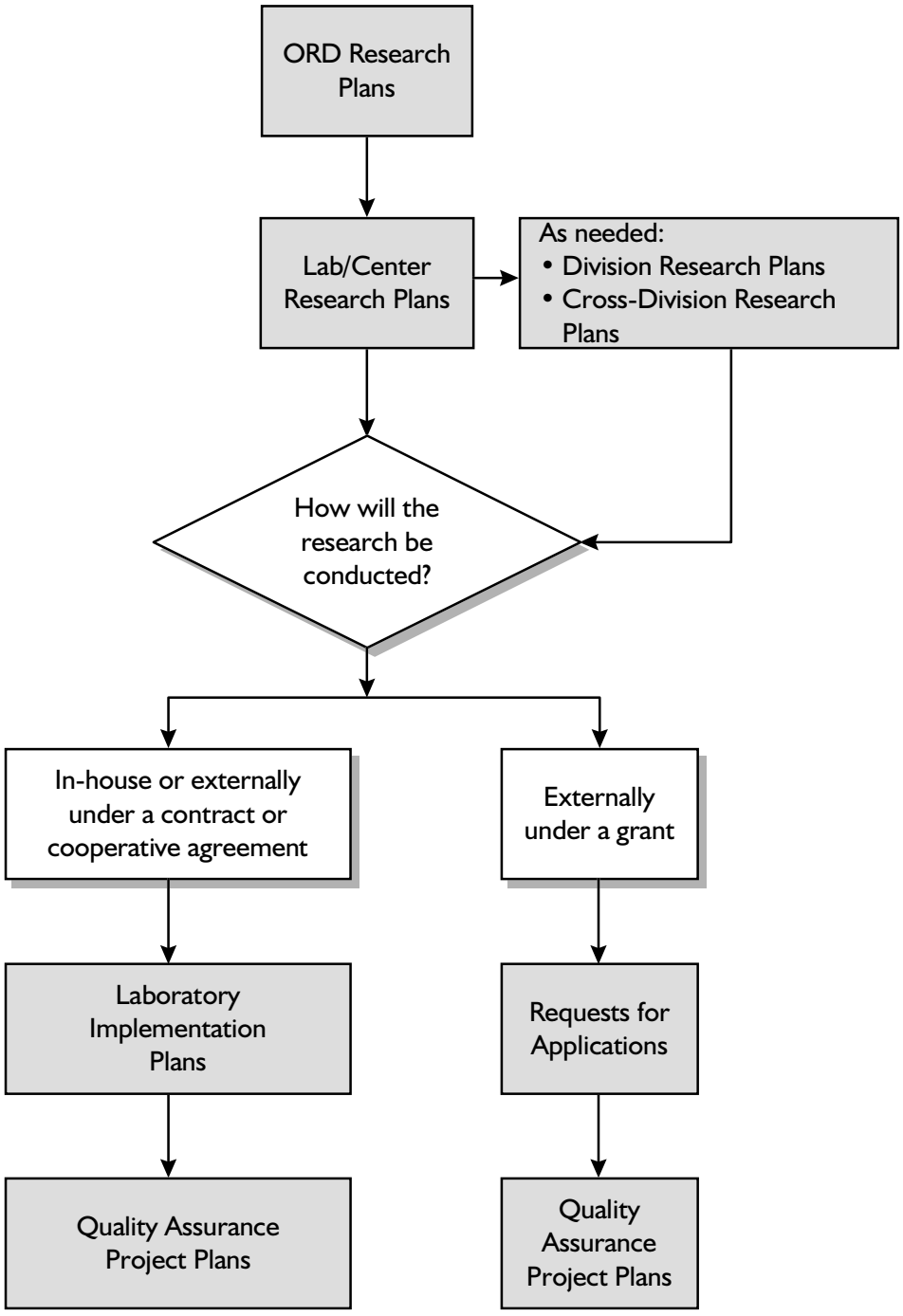
Attention to quality cannot be overstated, particularly in the context of IM planning. The existing EPA Quality System provides a framework and structure by which environmental programs produce results, including environmental data, that are of the type and quality needed and expected for their intended use. This is an Agency-wide, mandatory quality assurance program that includes all EPA intramural and extramural environmental data collection activities.[1]

Under Delegation of Authority-I-41, "Mandatory Quality Assurance Program," ORD is the focal point in the Agency for quality management system policy. The Assistant Administrator for Research and Development (AA/ORD) is designated as the Agency Senior Management Official for Quality Management. The Quality Assurance Division within ORD's NCERQA effects these responsibilities by developing quality assurance (QA) and quality control (QC) requirements and by overseeing Agency-wide implementation of the EPA Quality System.

Utilizing this Agency-wide system, ORD organizations implement approved quality management plans (QMPs) documenting their quality policy and quality system for applicable environmental programs. Quality systems include roles and responsibilities and processes for handling of documents and records and computer hardware and software.

[1]Environmental data include any information collected or produced from measurements, analyses, or models of environmental processes or conditions, or from experimental systems representing such processes and conditions, including results from laboratory analyses.

Figure 4-2. ORD Research Planning Documents



*Each organization has its own quality assurance system, with its policies, procedures and structure described in its Quality Management Plan.

Approved Quality Assurance Project Plans (QAPPs), or equivalent documents defined by the organization's QMP, are implemented for all applicable projects and tasks involving the collection or use of environmental data. QAPPs define and document the type and quality of data needed for the project and how specific QA and QC activities will be implemented and assessed during a particular project. Data management and handling of documentation and records are also elements defined in a QAPP. EPA QA/R-5, "EPA Requirements for Quality Assurance Project Plans for Environmental Data Operations," specify the QAPP data management requirements, which include:

- Describe the project data management scheme, tracing the path of the data from their generation in the field or laboratory to their final use or storage.
- Describe or reference the standard record-keeping procedures, document control system, and the approach used for data storage and retrieval on electronic media.
- Discuss the control mechanism for detecting and correcting errors and for preventing loss of data during data reduction (i.e., calculations), data reporting, and data entry to forms, reports, and databases.
- Identify and describe all data handling equipment and procedures to process, compile, and analyze data. This includes procedures for addressing data generated as part of the project as well as data from other sources.
- Include any required computer hardware and software and address any specific performance requirements for the hardware/software configuration used. Describe the procedures that will be followed to demonstrate acceptability of the hardware/software configuration required.

The guidance document "Technical Guidance G-5 Data Management Appendix" (EPA QA/G-5) provides additional guidance on the range of scientific and technical operations that involve generating, collecting, manipulating, and archiving environmental data.

Centralized Coordination of Data Management

The new ORD Science Information Management Coordination Board will give more detailed consideration to data collection, quality checking, and archival issues, including the possible role of a centralized data management function for ORD, such as that performed by NASA's National Space Science Data Center and NOAA's National Geophysical Data Center.

Awareness: Strategy Component 2

ORD has long utilized traditional outreach mechanisms (announcements, newsletters, publication lists, etc.) to make potential users aware of its publications. Under the awareness component of this IM strategy, ORD will add to its outreach portfolio electronic outreach via the Internet. Internet outreach will enable ORD to greatly expand:

- The number of people within and outside the Agency who are aware of ORD information resources.
- The information resources they are aware of—since the Internet offers the opportunity to readily create and update comprehensive, searchable directories of available information resources.[2]

Specific electronic outreach activities will include:

- Registering "ScienceNet"—ORD's home page and "front door" on the Internet—with search services that then will direct interested users to ORD's home page. (The "Access" section provides more information on ScienceNet.)
- Listing ScienceNet's Internet address in all printed ORD publications.
- Establishing links *from* ScienceNet to other appropriate Internet sites and working to establish links *to* ScienceNet from these sites.
- Establishing effective search capabilities and links within ScienceNet so that users who access ORD's home page will clearly know what types of ORD information resources they can access and how.

As described in the "Access" section, some information, such as draft reports and peer review documents, will be made available to selected EPA personnel and external partners under the ORD-wide Intranet or a special-purpose Intranet. For these items, ORD will conduct more targeted electronic outreach, via such mechanisms as electronic mail and brochures, to notify the appropriate audiences when the material becomes available.

[2] "Available information resources" can include a wide range of information, such as electronic and hardcopy information, ORD and external information, and data sets, databases and information products.

ORD also will take steps to increase awareness among ORD staff of how to use the Internet as a powerful research tool. Finally, ORD will conduct outreach to publicize to ORD clients and stakeholders the fact that ORD will be strategically managing its information and making it available to them.

Access: Strategy Component 3

ORD will continue to utilize some of the traditional mechanisms for providing access to information—for example, the Air and Radiation Docket, which enables individuals to inspect and make copies of documents—many of which have been in place for some time. Now, the Internet and other state-of-the-art information management technologies make it possible for ORD to:

- Greatly expand who has access to ORD information, how rapidly they can access it, and how much information they have access to.
- Tailor access to specific users or groups as appropriate for certain types of information. This tailoring is achieved by establishing limited access networks, called Intranets, that can be used to facilitate electronic information exchange among targeted groups or individuals both within and outside ORD.

Under this IM strategy, ORD will take full advantage of these possibilities by using the Internet (and specifically the World Wide Web) as the primary vehicle for enabling access to ORD data and information. Specifically, ORD will develop a coordinated information network that provides electronic access to its data and information through three mechanisms:

- An ORD ScienceNet (i.e., ORD's home page) that will be accessible by *all EPA employees and the public* via the Internet.
- A series of Science Intranets that restrict access to *specific users* who may include selected EPA staff, ORD staff, and authorized partners.
- Local and Wide Area Networks (LAN/WANs), which will be established for *specific ORD user groups*.

These components of the ORD information network are illustrated in Table 4-1 and described below.

ORD ScienceNet

As the "front door" to ORD's public scientific information, the ORD ScienceNet will be designed to provide a lively, exciting view of ORD's research. Through ScienceNet, users will have access to the published results of ORD research, which include printed documents, reports, and journal articles, as well as databases, data sets, models, and software applications (Figure 4-3).

Using a variety of tools (see the "Specialized Tools" section) available via the World Wide Web, users will be able to access, search, and download this information, as well as aggregate, manipulate, and analyze downloaded data. The ability to aggregate data may particularly aid environmental decision-makers, who often must synthesize large amounts of information during the decision-making process. For scientific data, users will be able to access metadata that describe the context and assumptions under which the data were collected (see the "Specialized Tools" section). Also, users will be able to link from the ScienceNet website to other information or scientific databases (such as STORET and Envirofacts) via hypertext links.

ScienceNet will supplement and, where appropriate, replace some of the more traditional delivery mechanisms (such as mailing hard copy) that ORD has historically used to disseminate information to the scientific community and the public. This will increase the effectiveness of ORD information dissemination, while lowering the cost.

One goal for ScienceNet is that it serve as a model and catalyst for creating two larger scientific information networks:

- An EPA-wide ScienceNet that would include the ORD ScienceNet.
- A broader U.S. ScienceNet developed collaboratively by EPA and other federal research agencies that would include EPA/ORD ScienceNet as one of many components.

ORD Intranets

ORD will develop a series of Intranets to share its internal scientific, management, and administrative information among EPA and ORD personnel. Intranets use the Internet to network computers together. Access is limited to those individuals whose computers are on the Intranet network. By using the Internet, Intranets circumvent the user incompatibility problems that may occur with

Table 4-1. ORD's Information Management Network

Network Component	What Information Will Be Available?	Who Will Have Access?
ORD ScienceNet	ORD information appropriate for release to the public. For example, users will be able to access published reports, data sets/databases, and grant & fellowship information.	All EPA employees and the public
ORD Intranets		
EPA-Wide Intranet	ORD information appropriate for use within the Agency. For example, EPA programs will have ready access to draft research plans and can comment electronically.	All EPA personnel
ORD-Wide Intranet	ORD information appropriate for distribution within ORD. ORD currently has an Intranet established for use by ORD employees.	ORD staff and authorized EPA and external partners
Special-Purpose Intranets	ORD information relating to specific research projects or groups within ORD.	Specific groups within ORD
ORD LAN/WAN Access	ORD electronic mail and specific administrative functions, such as OMIS.	All ORD staff

LANs—such as incompatibility between PCs and LAN software. The ORD Intranets will have two important purposes:

- First, they will provide access to materials that are either not relevant or appropriate for broader public access. These include:
  - Draft documents not yet cleared for public access.
  - Internal databases.
  - Administrative information in ORD's Management Information System needed for accountability, performance, and results tracking.
- Second, the ORD Intranets will provide a valuable mechanism for facilitating internal ORD processes that require a high degree of coordination—such as developing risk assessments, peer review, project collaboration, and coauthoring or reviewing draft documents. Temporary and permanent ORD Intranets can be established to accommodate these needs.

To fulfill these two purposes, ORD will establish two types of Intranets:

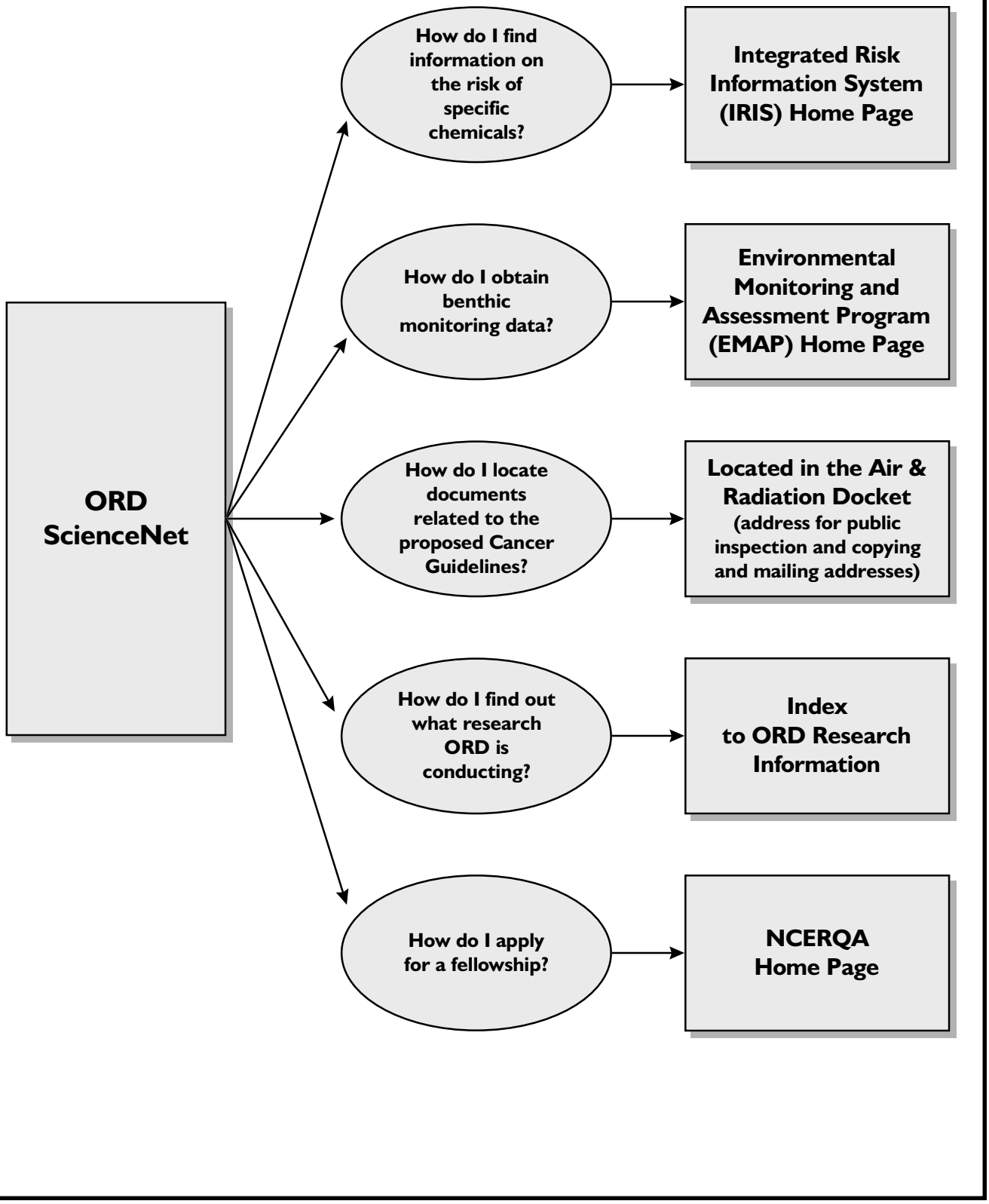
- A permanent ORD-Wide Intranet restricted to ORD staff and authorized EPA and external partners. This Intranet has already been established.
- Special-purpose Intranets to serve the particular communication and coordination needs of specific ORD groups, including established groups (such as a particular ORD Laboratory or Center) and ad hoc groups that exist for limited periods to conduct specific projects. Special-purpose Intranets will be created on an as-needed basis. They may be permanent or temporary.

The Intranets essentially will replace some of the current uses of ORD's LAN/WAN systems. To make the transition to Intranets as efficient, effective, and user-friendly as possible, the Intranet capabilities will build on and integrate the current LAN/WAN systems.

LAN/WAN

In recent years, ORD's LAN/WAN systems have served as the primary venue for internal ORD communication. LAN/WAN information transfer is primarily accomplished by a gateway server that transmits information via dedicated lines to servers

Figure 4-3. The ORD ScienceNet



at various ORD laboratories and/or via the Internet to targeted recipients.

As mentioned above, ORD will shift certain LAN/WAN uses to the Intranet where the latter approach offers important advantages. However, where it does not, ORD will continue to use, and will enhance as appropriate, its existing Local and Wide Area Networks. Future uses envisioned for the LAN/WAN systems include:

- Electronic mail and groupware such as Lotus Notes.
- Various scientific and administrative software applications such as GIS, visualization, and statistical tools that are not yet Internet enabled.
- Certain administrative and management functions, such as OMIS.

Status of Access Components

Internet Operating Plan
Development of ORD's electronic information network is well underway. ORD's National Risk Management Research Laboratory in Cincinnati, Ohio, is finalizing an Internet Operating Plan that defines the resources, policy oversight, and services ORD will need to achieve its electronic information access goals—including development of the ORD ScienceNet and the ORD Intranets. The Internet Operating Plan will form the basis for ensuring that the ORD ScienceNet and Intranets are state-of-the-art information management systems supported by up-to-date telecommunication and other transmission equipment.

Development of ScienceNet
In addition to the operating plan discussed above, ORD will inventory current electronic products already created by ORD staff and determine which should reside on or be linked to on the ORD ScienceNet. Candidates for ScienceNet include ORD-relevant databases (an inventory of these databases is provided as Appendix C) and the home pages that already have been created by many ORD groups.

Status of OMIS Implementation
This action also is well underway. By the end of FY97, OMIS will hold complete and organizationally consistent information on ORD research projects and tasks, commitments, obligations, and expenditures; grants and contracts; and training plans. This information will be updated on an annual basis. Currently, ORD staff access OMIS via the ORD's LAN/WAN system. One action item for OMIS is to decide whether to shift OMIS access to the ORD Intranet. This decision will be made after evaluating the potential benefits of such a change.

LAN/WAN Maintenance
LAN/WAN systems already are well established at ORD. ORD will continue and, where appropriate, enhance their use to support electronic mail, some administrative and management functions (e.g., OMIS), and limited file sharing (O-drive) capabilities.

Incentives for Sharing Data

In addition to expanding access to information through better use of information technologies, ORD will promote an organizational culture of sharing data. Examples of how ORD will more strongly encourage this include:

- Establishing an internal reward system for researchers who publish data sets.
- Developing a policy that affords researchers a certain time period for data collection, analysis, and publication, after which the data will be generally available to others.
- Requiring that milestones for data delivery be included in ORD Research Plans.
- Requiring that an external grant-holder submit data collected during the research project, along with appropriate documentation, for subsequent use by other researchers.

As part of its implementation planning, the ORD Science Information Management Coordination Board will expand on these four examples and fully address how ORD should establish mandates and incentives for individual researchers and programs to make their data and information generally available.

Usability: Strategy Component 4

Usability—making sure that the access system as well as the data and information provided by the system are usable to internal and external users—is vital to the success of this IM strategy. This is true for ORD's traditional access mechanisms as well as the more recent electronic ones. To optimize usability ORD must:

- Design the information access mechanism/network to meet users needs (described in Section 3) in as complete and efficient as manner as possible.

- Provide the infrastructure and support to enable ORD staff to place important information on the system in a timely manner, manage the information, and effectively utilize the hardware and software components of the system.
- Ensure that adequate documentation can be obtained so that ORD data and information that is located and retrieved is also usable.

Though a robust infrastructure will require investment, the potential returns are substantial. As discussed in Section 3, a well-functioning information management system will greatly enhance ORD's internal functions and external profile.

Usability involves a number of components:

- Security and access control.
- Training.
- Standards.
- User support services.
- Navigational aids.
- Policies.
- Specialized tools.
- Specialized information resources.

These components are discussed below. The following text box on "Cars and Information Systems" provides insight into why they are important.

Security and Access Control

Security measures guard against hackers who seek to tamper with information and damage information systems. *Access controls* ensure that only authorized users can create, read, modify, and delete information. Access control is particularly important for sensitive information, such as personnel and budget data.

A well-managed and adequately staffed security and access control program is essential to safeguard the integrity of the information and data that will be housed in ORD's system. Without these safeguards, ORD would be vulnerable to accidental and deliberate disruptions to its ability to use its information. Such disruptions can be very expensive.

Using guidance from EPA's Office of Information Resources Management, ORD will perform an in-depth security risk assessment that covers all ORD data centers. Based on the outcome of this assessment, ORD will take the required steps to ensure the integrity of its data collection. These steps will include:

Of Cars and Information Systems: Why Usability is Important

Driving a car provides a useful analogy for why usability is vital to a successful information system. Just as drivers need far more than cars to reach their destinations, information consumers need far more than technology (hardware and software) alone to successfully access and use information. In both cases, success depends on a robust infrastructure:

- **Security and access control/permission.** Automobile drivers need "permissions" to drive—insurance, registration, driver's licenses, etc. Computer users need analogous permissions, such as valid software licenses and authorizations to create, read, modify, and delete appropriate categories of information.
- **Training.** Computer users need their own version of "driver's ed": a well-conceived training program that teaches them how to use their information technology efficiently and take advantage of new software applications.
- **Standards and policies.** Standards establishing rules of conduct are crucial for safe and pleasant driving. Likewise, baseline standards and policies for information technology are important to make users' experiences predictable, productive, and even pleasant. For example, ORD users would benefit from clear and concise policies about sharing data and information, security practices, and training.
- **User support services.** Drivers need support services, including service stations and other facilities, to help them when they require assistance. Similarly, computer users need user support services to help them when their computer "breaks down" or when they get "lost" in cyberspace.
- **Navigational aids.** Drivers need navigational systems, including signs, signals, and maps, to help them reach their destination efficiently and safely. Information users need equivalent aids to help them navigate around information resources. These may include, for example, metadata that explain how information was developed and aid understanding of how it can be used and standard mechanisms to help users understand where they are and how to "go home."

- Developing security policies, procedures, and standards that leverage security documentation and tools already used by the Agency.
- Investigating tools that can aid ORD in meeting its security objectives in a client/server and Internet environment.

Training

ORD traditionally has relied on Agency-sponsored training on discrete topics. Training sessions have tended to use conventional formats, such as topic experts training relatively small groups. This approach has limitations. In particular, it tends to create a situation where the educational needs of ORD staff are met sporadically and inconsistently.

For information management, ORD needs a comprehensive, coordinated, and strategic approach to training to ensure that:

- *All* staff receive training in basic skill areas necessary for information use.
- More specialized training in a timely manner is available to those who need it.

To achieve these goals, ORD will expand its portfolio of training approaches to include computer-based training, video conferencing, and outsourcing training. Also, ORD will develop and deliver to ORD staff:

- An ongoing training and certification program for core competencies in the basic ORD information environment.
- As-needed training for higher levels of expertise in key areas such as GIS and risk assessment.

Standards

ORD will search out and adopt appropriate information management standards at the highest level applicable, for example, those developed and promulgated by ISO, ANSI, and NIST. When appropriate standards do not exist, ORD will develop them to ensure that its information is usable. For scientific data, standards will include descriptive information about the data (metadata). For example, ORD staff will be better able to utilize archived scientific data for future research if they have access to metadata they can use to evaluate the quality, relevance, and practical utility of the archived data (see also the "Specialized Tools" section).

Currently, descriptive elements and quality assurance codes established in 1987 under the *EPA Standards for the Electronic Transmission of Laboratory Measurement Results* (EPA Order 2180.2) are the only scientific metadata requirements available. These requirements will be updated, expanded to cover field data, and made consistent with ORD record retention schedules.

The Agency has a number of other standards that address various aspects of information management:

- Uniform Rulemaking Docket Manual.
- Facility Identification Standard.
- EPA Hardware and Software Standards.
- Security Audit Standards for LANs and PCs (i.e., Enterprise Security Manager).

These standards will be updated as needed to respond to the additional needs of ORD's new information management approach.

User Support Services

ORD users will need support services to help them when they have questions or encounter difficulties. Since user needs will vary considerably, ORD will develop a broad spectrum of service response mechanisms as it implements this strategic plan. Examples of the types of support services ORD may provide to its users include:

- *An electronic "HELP" menu* to assist users by topic area—much like the help menu in a software package. The help menu will be included as an integral component of directories, catalogs, and dictionaries (see the "Specialized Tools" section).
- *An electronic mail and/or telephone hotline*, staffed by IM specialists to aid users in locating, accessing, collecting, and analyzing data. When developing the hotline, ORD will research current EPA hotlines to identify appropriate models for the IM hotline.
- *A published (electronic) inventory of ORD IM specialists* and their respective areas of expertise, as well as extramural data processing and analysis mechanisms available to Agency users.

Navigational Aids

ORD information usability will benefit from a navigational framework that includes standard mechanisms to help users understand where they are in information space, how to "go home," how to get "help," and so on. For key information areas,

such as risk assessment, the equivalent of "tour guides" may prove very useful for helping ORD researchers understand how they can apply these information resources in their work.

ORD will integrate into ORD ScienceNet a navigational framework that includes standard web search engines and the ability to easily browse ORD's information (Figure 4-3). The ScienceNet also will contain a "highlights" section that will focus on novel applications of IM technologies in environmental research, many of which will come from ORD's own research programs.

In addition, ORD will develop topic search agents that ORD staff can use to automatically search the Internet and locate information pertinent to their research.

Policies

Appropriate policies are crucial for fostering information use. ORD may want to establish a policy encouraging the sharing and publication of information, based on a reward system for researchers who publish data sets. Another simple example involves "credit and control." Information suppliers want credit for the information they supply, typically in the form of appropriate citations and acknowledgments comparable to the norm in scientific publications. Similarly, they want to feel they have some control over how the information they supply is used. These issues can be addressed by simple policies that promote adequate documentation of:

- Data sources and contacts.
- The purpose for which the data were collected.
- The documented quality of the data.
- The fact that the data were validated as appropriate for the intended use.

The federal government, EPA and ORD currently have a number of policy and guidance materials for using ORD information systems, including:

- IRM Policy Manual (2100).
- IRM Privacy Act Manual (2190).
- System Design and Development Guidance.
- EPA Operation and Maintenance Manual.
- Records Management Manual.
- Enterprise Technology Services Division (OIRM/ETSD) Operational Directives.

While these policies are largely sufficient in their current form, not everyone in ORD is familiar with them, and overlap and conflict among all of these policies undoubtedly exists. ORD will first reconcile differences and fill gaps in the policy and guidance materials and then widely disseminate them among its staff through training programs. It will provide regular oversight and support to ensure that these policies are implemented.

Specialized Tools

Some ORD information users will need specialized tools that synthesize, analyze, and display data and information to generate new information and knowledge. Key tool categories include:

- Geographic Information Systems.
- Statistical tools.
- Computer models of all sorts (including those needed for risk assessment and risk management).
- Visualization systems that help users visualize what data mean.

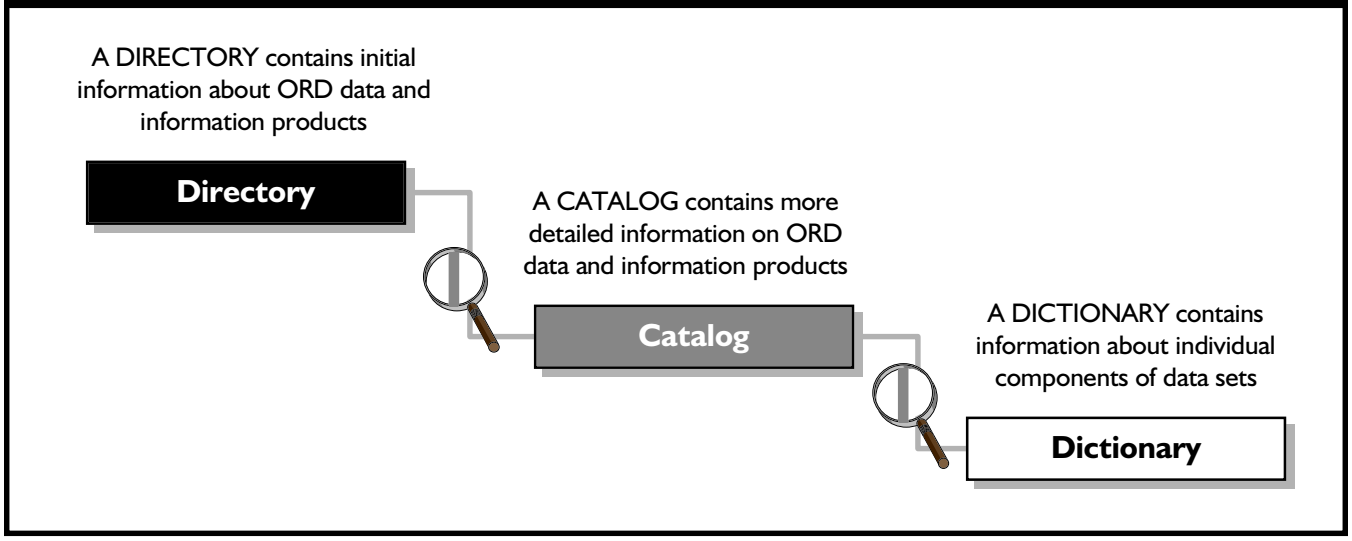
ORD will identify priority needs for these tools and make sure they are available to ORD and external users who need them.

One particularly noteworthy tool that ORD will provide to make its science data more usable is a searchable "directory/catalog/dictionary" database system that will be available on the ORD ScienceNet. From the ScienceNet, this database tool will enable any user to rapidly search ORD's science data collections, identify data relevant to the user's particular need, understand its constraints, and determine whether it is useful for the user's purpose. The directory, catalog, and dictionary components of this system are illustrated in Figure 4-4 and described below.

Directory
The ORD *directory* will consist of a database containing information about the scientific information products produced by or relevant to the ORD mission. Types of information products include data sets, databases, models, projects, analytical products, and documents.

The ORD data directory will provide sufficient information to enable ScienceNet users to determine which of the available ORD science data products (or objects) may be relevant to their needs. For each product (or object), the directory will provide information on:

Figure 4-4. Components of a Searchable Database (the DCD System) that Will Describe the Information Available on ORD ScienceNet



- Why the data were collected.
- Who collected the data.
- Where the data were collected.
- How they were collected.
- When they were collected.
- The format of the data set.
- What fields the data set includes.
- Other information as appropriate.

Catalog
The ORD catalog provides an evaluation mechanism to assist consumers in deciding if a product, such as a model or data set, will be useful in meeting their needs. The ORD catalog will provide detailed information on:

- The general purpose for which the data were collected.
- The originator of the data.
- Sampling and laboratory methods.
- Descriptions of the data and any manipulations or transformations of the data.
- Related quality control/quality assurance measurements.
- Procedures necessary for data access.
- References to publications that use the data set.

The level of detail provided in the ORD catalog will allow most data users to decide whether the data are adequate for the intended use.

Dictionary
A dictionary provides a further level of detail for science data users: descriptive information about *individual fields or attributes* in data sets and databases. This is in contrast to the directory and catalog that relate information to entire data sets (as well as other types of information products).

Status of DCD System Implementation
ORD's National Center for Environmental Assessment currently is assisting EPA Region 10, the Office of Water Surf Your Watershed program, OIRM's Envirofacts, and ORD NERL in developing and implementing a directory, catalog, and dictionary (DCD) system called the Environmental Information Management System (EIMS) to manage environmental data relevant to assessment activities. These systems are currently operational.

The EIMS is being evaluated for use as an Agency-wide metadata management system. The ORD EIMS will be available to users through the World Wide Web and will have the following capabilities:

- The *directory* and catalog have both update and browse capability.
- The *catalog* will provide detailed information on the quality of data sets, the methods employed to create the data sets, and other information consumers can use to determine the appropriateness of the data for their purposes.

- The *dictionary* will provide information for all fields within the ORD data set and database collection. It will be fully integrated and nonredundant with the directory and catalog.

To be successful, ORD's DCD system must be supported by a strong data administration program that establishes:

- What documentation ORD staff should provide to the system.
- How they should provide it.
- Where to provide it.
- When to provide it.
- Who is responsible for providing it.
- Why it should be provided.

ORD will develop data administration policies, standards, and procedures that address these questions and establish consistent ground rules for both contributors and consumers of ORD metadata to follow when populating and using the DCD system.

Specialized Information Resources

Some ORD users will need specialized information resources that come from suppliers outside of ORD and, often, outside of EPA. A good example is remote sensing data. For research themes like ecosystem protection, remote sensing data can provide tremendously improved insights and synergy with other data.

ORD will survey its staff to determine which specialized data resources are needed. Then ORD will work with the providers of these data (such as the National Aeronautics and Space Administration and the National Oceanic and Atmospheric Administration) and use interagency mechanisms (such as the Committee on Environment and Natural Resources and its subcommittees and task forces) to make its needs known. Finally, ORD will prepare a compendium of specialized data resources with information about what they contain and how to access them and distribute the compendium to ORD staff.

ORD Science Information Management Coordination Board

The newly formed ORD Science Information Management Coordination Board, as described in Section 1, will be responsible for developing implementation plans for this strategy and for recommending an appropriate FY 1998 and 1999 budget to support high-priority ORD science information systems infrastructure.

Timing

The need for IM capabilities within ORD is so great that many IM activities or systems already are being implemented by individuals or groups within ORD. This is inherently inefficient because the time and resources to establish these systems independently is far greater than a coordinated effort. Also, the end results often are incompatible with other ORD systems or, at the very least, far less accessible and useful to others than they could be. For example, several large ORD initiatives, such as the Environmental Monitoring and Assessment Program (EMAP), have an urgent need for a coordinated data management system. If ORD does not implement a

Range of IM Needs Among ORD Laboratories and Centers

- The primary mission of ORD's *three Laboratories* is to conduct research. This generally involves collecting and analyzing large amounts of data. The initial users of these data are the ORD and cooperating scientists conducting the research.
- By contrast, the mission of ORD's *National Center for Environmental Assessment (NCEA)* is to provide support for ORD and stakeholder use of risk assessment and conduct risk assessments as a secondary data user. NCEA typically synthesizes data generated by ORD Laboratory research, conducts secondary analysis of this data (e.g., uses it to perform a risk assessment), and disseminates the results to its customers and stakeholders.
- With its grants and fellowship program management, quality assurance, and peer review responsibilities, ORD's *National Center for Environmental Research and Quality Assurance (NCERQA)* has yet another distinct set of information management needs to be able to provide adequate management for ORD's extramural grant and fellowship programs and ensure that the best approaches to environmental data quality and documentation are incorporated throughout ORD.

coordinated IM strategy in the very near future, this program and others likely will be forced to act on their own to stay on schedule with the project goals and milestones.

In general, the more time that passes before an ORD-wide IM system can be created, the greater will be the inefficiencies and the greater will be the cost to integrate existing IM components into a unified system.

External Coordination

Many organizations within and outside EPA are planning for and implementing information management systems for a variety of purposes. The Board will need to be proactive in contacting and communicating with other groups in EPA, with ORD's stakeholders, and with our partners to obtain feedback on this plan before and during implementation and to stay abreast of and coordinate with initiatives, plans, and programs related to our IM efforts. This coordination will be accomplished in part by the outreach activities discussed earlier in this section, which will serve to make ORD's stakeholders and partners within and outside the Agency aware of ORD's IM activities. For example, in the near term, the Board will need to coordinate with EPA's efforts to implement a consolidated planning, budgeting, and accountability system. Also, as described in Section 1, the Board should coordinate with EPA in its efforts to comply with the Information Technology Management Reform Act.

Successful implementation of ORD's IM Strategic Plan will also require substantive input and involvement by all ORD Laboratories, Centers, and Offices. Therefore, during the execution of this plan over the coming years, an effective staffing structure with sufficient management support and resources to operate effectively and responsively will be crucial to ensuring ORD-wide coordination and participation. This is particularly important because ORD is a large organization distributed across the United States (Appendix D), which heightens both the need for and challenges of ORD-wide IM management and coordination.

Each of ORD's five Laboratories and Centers and three Offices currently has its own IM staff (see Table 4-2) responsible for meeting the internal IM needs of its own researchers, managers, and administrative personnel in a way that aligns with its mission. What is currently lacking is a unifying mechanism providing cohesion and overarching coordination among these staff on an organizational level. It is, therefore, recommended that the Board initially look not only at resource needs, but at a staffing structure that will enable ORD's IM staff to more effectively coordinate IM activities across ORD. The challenge for ORD will be to increase coordination, training opportunities, and other support to our scientists and engineers, while avoiding unnecessary bureaucracy.

Table 4-2. Current ORD Information Management Staff*

Organization	Site	Computer Specialist/ Engineer/ Assistant	Information Management Specialist/ Assistant	Technical Information Specialist/ Assistant	Computer Graphic/Visual Information Specialist	Management/ Program Analyst	Scientist, Engineer, or Mathematician
HQ	Washington, DC	4				3	
	RTP, NC	1					
NCEA	Washington, DC		6			1	
	Cincinnati, OH	1		1			
	RTP, NC			1			
NCERQA	Washington, DC		1				1
NERL	RTP, NC	6					
	Cincinnati, OH	1					1
	Las Vegas, NV	2	1		1		
	Athens, GA	1	1				
NHEERL	RTP, NC	10	2		1	1	
	Narragansett, RI	1		1		1	
	Corvallis, OR	1	1				
	Duluth, MN	1				1	
	Gulf Breeze, FL						
NRMRL	Cincinnati, OH	7			1		2
	Ada, OK	1					
	RTP, NC	1					

ORGANIZATIONAL ABBREVIATIONS:

HQ - Headquarters
NCEA - National Center for Environmental Assessment
NCERQA - National Center for Environmental Research and Quality Assurance
NERL - National Exposure Research Laboratory
NHEERL - National Health and Environmental Effects Research Laboratory
NRMRL - National Risk Management Research Laboratory

* Many of these individuals perform duties outside the strict definition of information management, such as responding to Freedom of Information Act requests, producing risk assessments and other documents, managing contracts, and handling review and clearance of products.

Appendix A:
The Information Technology Management Reform Act of 1996 (ITMRA)

A fundamental change in the way government agencies perform the work of information management was signed into law in February 1996. The ITMRA repeals large portions of the Brooks Act and establishes a new statutory scheme for information technology management and acquisition. The ITMRA formally establishes the position of Chief Information Officer (CIO). The CIO must put in place a process to meet new requirements for Information Technology Capital Planning and Investment Control.

The CIO will consider mission accomplishment first in approving major systems, overarching policies, agency acquisition strategies and budget. Policy issues will be resolved *before* major system automating efforts begin.

These new requirements raise several challenges for the Agency to meet the goal of the Act, namely to maximize the value of our investments in information technology while minimizing risks:

- Work Process Reengineering as a prerequisite before investing in automation, along with determining whether processes should even be automated or should be performed by entities other than EPA. If we are going to continue to do the work, we must insure that we automate the best approach to the work.
- Return on Investment, including explicit measures of cost and benefits to be realized, as well as cost avoidance, and least costing option. In many cases, the risk on investment will not necessarily be all quantitative but qualitative as well. Investments will be ranked and prioritized that meet Agency thresholds ($25 million/life cycle, $5 million /year). Questions of aggregation of technology and the application of granularity must be answered.
- Risk Management, including explicit criteria for evaluating risk, leading to relative rankings of the associated risks across IT projects considered for investment.
- Performance Measurement, leading to tracking of systems as being on time and on budget plus meeting requirements objectives. The legislation specifies that the Agency must achieve a 5% reduction in costs due to automation.
- Portfolio Management, including dealing with systems in several stages of life cycle, risk, and cost. Major investments must come before an investment board (i.e., EPA's Executive Steering Committee) with an accompanying Business Case Analysis that will summarize the mission benefit as well as the newly introduced Return on Investment and Risk measures.
- We must define investment and thresholds which includes the ongoing costs of operating systems, all life cycle investment, and at the appropriate level of management, rank and prioritize these systems based upon mission accomplishment.

EPA must merge the existing and new processes to the greatest extent possible to create the necessary Capital Planning and Investment Control process.

Action

ORD must approach each information technology investment as a business case and apply rigorous cost benefit analyses. An Agency-wide architectural

process has begun that will drive home the need for a cultural change. The fundamental change required is that no added costs or time delays to systems may be incurred. We must begin now to:

1. Invest human resources in the Strategic Information Technology Architecture Plan (SITAP) process.
2. Participate in the ITMRA Capital Planning and Investment Control Strategic Project Committee to define the Agency process for capital planning and investment control, determine best practices, and implement the new method of determining information technology investments.

Appendix B:
ORD's Strategic Research Planning Process

As described in Section 1, ORD has instituted a strategic planning process to determine research priorities based on risk assessment and risk management principles. This process was finalized in 1996 and published in the *Strategic Plan for the Office of Research and Development*. This strategic plan established the framework by which ORD determines the highest priority research to which intramural and extramural research efforts are directed.

The risk-based research planning process involves soliciting stakeholder input, identifying potential research topics, evaluating these topics to select the most appropriate ones for ORD's research agenda, and then defining specific research needs for the selected topics. Once ORD has identified specific research needs, plans are developed that define long-term directions as well as the immediate research that will be carried out during the next few years. The ORD laboratories then prepare plans that describe how they will implement the intramural component of the research, and complementary solicitations are prepared for extramural research that will be carried out by universities and other research institutions. Research is conducted based on these plans; the research products are delivered to stakeholders and provide input into the next strategic planning cycle. Steps in the research planning process include:

Obtaining EPA and Stakeholder Input. The first step in ORD's research planning process involves seeking input from all parts of EPA (including ORD and the EPA Program and Regional Offices), as well as external stakeholders, to identify the most important and relevant areas for our research efforts. At this stage, we also consider the status and results of our recent research activities. Based on this information, we identify potential research topics.

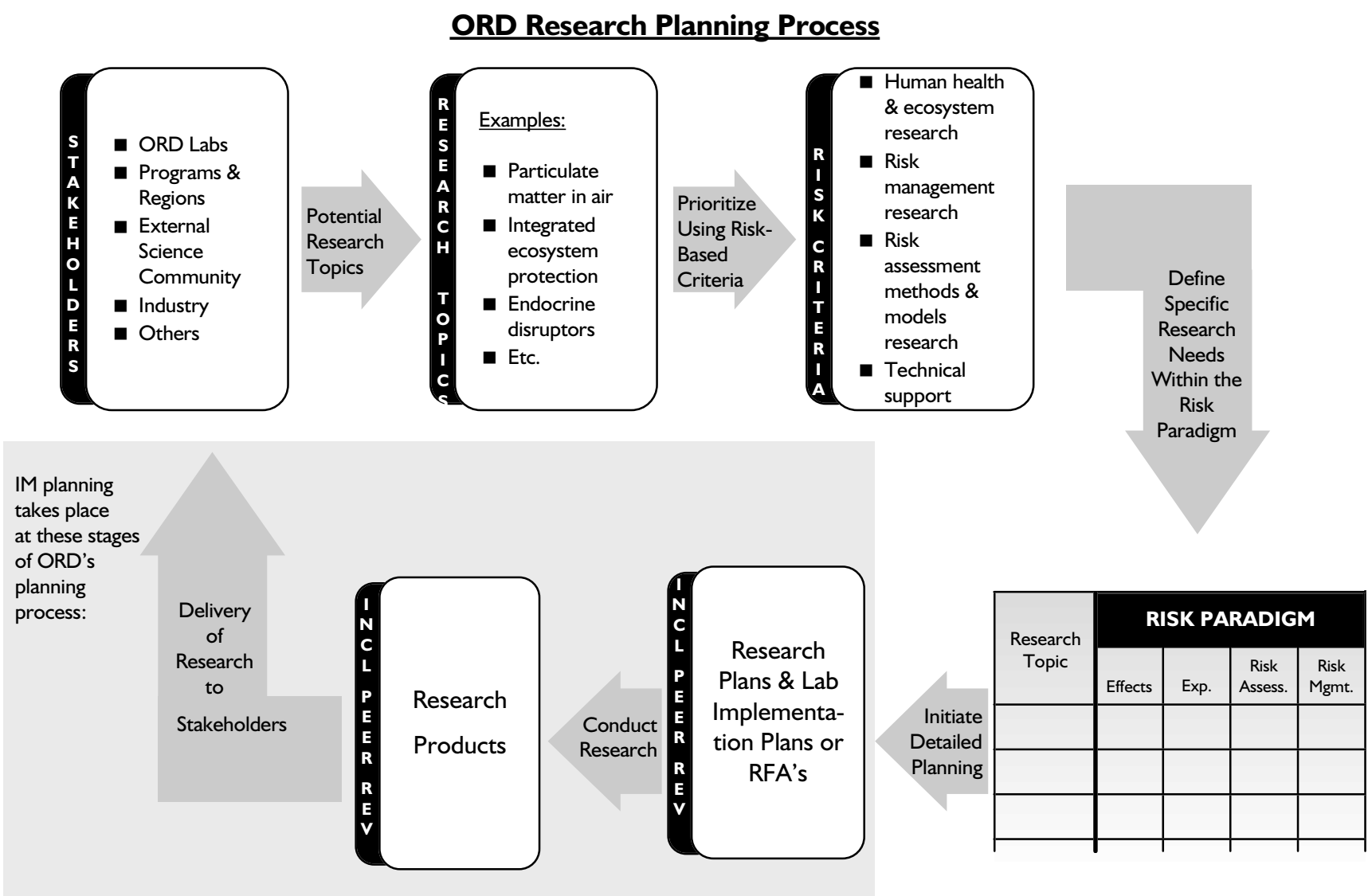
Selecting Research Topics. We narrow the pool of potential research topics by retaining only those that are either within ORD's mission or clearly mandated. To the mission-related topics, we apply a series of human health, ecological health, methods/models, and risk management criteria to compare the topics according to their potential to support effective risk reduction and/or produce broadly applicable results. We then further narrow the pool of topics by retaining only those where ORD can make a *significant* contribution to environmental science. Based on this analysis, we prioritize the research topics.

Determining Specific Research Needs. For each selected topic, we then determine what the specific research needs are within each component of the risk paradigm (effects, exposure, risk assessment, risk management). As appropriate, we may develop strategic plans for broad, overarching topic areas to guide more detailed planning.

Detailed Planning. The next step involves translating these needs into a research program. At this stage, we develop detailed research plans for each project. We also develop operating and laboratory implementation plans for each project to be conducted intramurally or under a contract or cooperative agreement. We develop Requests for Applications for projects to be conducted under grants.

Conducting the Research. Once planning is completed, the research is conducted intramurally or extramurally. The research products and results provide input into future planning efforts.

Relationship of IM Planning to ORD's Research Planning Process



Recommendations for Integrating IM Planning Into the Development of ORD Research Planning Documents

Planning Document	Description	Contents	Current Guidance on IM Planning	Recommendations for IM Planning
ORD Research Plans	Delineate overarching research strategies in the broad topic areas identified for ORD research (e.g., ecosystem protection, drinking water). Used by ORD and its stakeholders, including peer review panels, EPA Program and Regional Offices, and external stakeholders.	**ORD Research Plans describe:** How and why the topic was selected for ORD research (i.e., what criteria were applied, who the research clients are, and how the research will meet their needs). Specific research efforts and major products.	ORD *Guidelines for Research Plans* (October 13, 1995) do not require that Research Plans address IM planning.	ORDs *Guidelines for Research Plans* should formally incorporate a requirement that ORD Research Plans describe a general approach for managing the data and information that will be generated by the research.
Laboratory, Center Research Plans	Prepared by ORD 's National Laboratories and Centers as planning tools to identify, prioritize, and justify their overall programs.	**Lab/Center Research Plans:** Broadly define major research areas. Describe anticipated products or accomplishments and how the research will contribute to addressing /reolving Agency problems. Provide information on the current research program and future directions.	There currently are no formal guidelines for preparing these plans.	These research plans should describe a general approach for managing the data and information that will be generated by the research project.
Divisional Research Plans	Prepared by individual division within an ORD Laboratory or Center, these research plans define, prioritize, and justify the research program to be conducted by the division.	**Divisional Research Plans:** Provide detailed information on specific research areas, disciplines, or themes. Describe the relationship of the research to specific ORD research strategies. Describe the anticipated products/accomplishments and how they will contribute to addressing/resolving identified data gaps.	There currently are no formal guidelines for preparing these plans.	These research plans should describe a general approach for managing the data and information that will be generated by the research project.
Cross-Divisional Research Plans	Prepared by cross-divisional research teams, these plans describe research to be conducted collaboratively across divisions within or across ORD Laboratories and Centers.	**Cross-Divisional Research Plans describe the:** Problem and its importance. Research approaches. Anticipated research products/accomplishments.	There currently are no formal guidelines for preparing these plans.	These research plans should describe a general approach for managing the data and information that will be generated by the research project.
Laboratory Implementation Plans (LIPS)	Prepared by ORD Laboratories for all research to be conducted in-house or under a contract or cooperative agreement. Used by ORD managers and scientists to track and prioritize products (reports, data and metadata, publications) that present the science data and information gathered by the research. All LIP information resides in the LIP module of the ORD Management Information System (OMIS).	**Laboratory Implementation Plans:** Provide detailed descriptions of the specific research projects and tasks to be conducted. Include data on extramural resources and FTEs (full-time equivalents), products to be generated, and the acquisition or assistance mechanism for the research. Show how the resources will be used to implement the research programs and objectives. Divides the research into a set of discrete tasks, each assigned a unique task number. Specify what mechanism will be used to conduct the task (i.e., intramural, contract, cooperative agreement) and what products (i.e., research plans, reports, etc.) will be generated under the task.	There currently are no formal guidelines for preparing these plans.	LIPs should describe the specific tasks and resources needed to effectively manage the data and information that will be generated by the research.

Recommendations for Integrating IM Planning Into the Development of ORD Research Planning Documents

Planning Document	Description	Contents	Current Guidance on IM Planning	Recommendations for IM Planning
Request for Applications (RFAs)	Prepared under ORD's Science to Achieve Results (STAR) Program to solicit investigator-initiated grant applications from universities and other external research institutions in areas of special interest to EPA's mission.	**RFAs specify the requirements and provisions for the research grant. As appropriate, RFAs may also:** Require the investigator to summarize in the proposal the plans for data management. Make these plans one of the selection criteria. Require that the investigator submit data collected during the project, along with appropriate documentation, to an EPA data archive for subsequent distribution and use by other investigators. Require the investigator to describe in the proposal how the data will be preserved and made available to others for future research studies.	There currently are no formal guidelines for preparing RFAs.	When preparing an RFA, ORD should consider the need for IM planning and incorporate requirements for IM planning into RFAs as appropriate on a case-by-case basis. IM plans submitted in the grant proposal would then be peer-reviewed along with the rest of the proposal.
Quality Assurance Project Plans (QAPPs)	Prepared for all applicable projects and tasks involving collection or use of environmental data, to define QA/QC activities and records.	**Quality Assurance Project Plans (QAPPs):** Define and document the type and quality of the data needed to meet the intended use of the project. Describe how specific quality assurance and quality control activities will be implemented, assessed and documented during a particular project.	EPA QA/R-2 and R-5 require that data management and handling of documentation and records be defined in a QAPP.	Particular attention should be focused on documenting requirements of scientific data users, involving them in user acceptance testing and documenting test results as well as recordkeeping for baseline system and software applications and data and subsequent change management.

Appendix C:
Inventory of ORD-Related Databases

[Compiled by Sidney Draggan, February 27, 1996]

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks
Air and Radiation	*Clean Air Status and Trends Network (CASTNet)	Deposition network quantifying status and trends of air emissions, depositions and air quality [EPA 51/Total 235].	Office of Research and Development, NERL	Rick Linthurst 919.541.4909 Jay Messer 919.541.1425	EPA/600/A-94/194	
	*GLAD	Deposition of atmospheric trace elements and nutrients in precipitation to the Great Lakes [EPA 7].				
	*IADN	Identify persistent and bioaccumulative toxic pollutants, estimate their total deposition, and determine sources and relative loadings to the Great Lakes [5].				
	*IMPROVE	Establish current background visibility levels; identify chemical species and emissions sources responsible for existing manmade visibility impairment and document long-term spatial and temporal trends [53].				
	*NAMS/SLAMS/SPM	Air quality data for areas with high pollutant concentrations and high population exposure [507].				
	*PAMS	Comprehensive and representative data on ozone pollution non-attainment areas [144].				
	Aerometric Information Retrieval System (AIRS)	National repository for airborne pollution in the United States and various World Health Organization (WHO) member countries.	Office of Air Quality Planning and Standards; Information Transfer and Program Integration Division	1-800-333-7909	http://www.epa.gov/airs/airs.html	
	(AQS) Air Quality Subsystem	Contains measurements of air pollutants and meteorological data from about 10,000 monitoring stations operated by EPA, state and local agencies, and WHO member countries.				
	(AFS) AIRS Facility Subsystem	Contains data for nearly 150,000 air pollution point sources monitored by the U.S. EPA and/or state and local air regulatory agencies.				
	(GCS) Geographic, Common, and Maintenance Subsystem	Contains reference data shared by the AQS, AFS, and AG subsystems.				
	(AG) AIRS Graphics	Integrates data from multiple AIRS subsystems into maps and charts that show patterns, trends, and anomalies in air pollution data.				
	(AE) AIRS Executive	An IBM PC program that contains a select subset of data extracted from the AIRS database.				
	Support Center for Regulatory Air Models (SCRAM)	Provides regulatory air quality model computer code, meteorological data, documentation, and modeling guidance.	Office of Air Quality Planning and Standards			

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks	
	Clearinghouse for Inventories/Emission Factors (CHIEF)	Contains information on air emission inventories and emission factors, provides access to tools for estimating emissions of air pollutants and performing air emission inventories for both criteria and toxic pollutants.	Office of Air Quality Planning and Standards				
	UV-B Monitoring	Information supporting international program for ground-level measurement of UV-B [7].					
	Southern Oxidants Study (SOS)	Photochemical; oxidants-related ambient and emissions management data collected in the Southeastern U.S.	Office of Research and Development (ORD), NERL-Atmospheric Sciences Modelling Division	Basil Dimitriades 919.541.2706			
	An Operation System for Predicting the Population Health Effects from Disposal of Radioactive Waste by Shallow Trenches (PRESTO-POP)	Designed to calculate the population health effects resulting from the disposal of low-level radioactive waste in shallow trenches; employs simple radionuclide transport and exposure submodels.	Office of Research and Development	Hung, Cheng-Yeng 202.233.9204			Model
	Indoor Air Quality in Large Office Buildings	Contains ORD and OAR data; characterizes the temporal and spatial variability of important physical, environmental, and comfort parameters influencing occupant perceptions of indoor air quality. A standardized investigative protocol was employed to uniformly collect and report indoor air quality data from randomly selected office work environments in randomly selected public and private office buildings.	Office of Air and Radiation	Susan Womble ORIA/OAR 202.233.9057 Ross Highsmith NERL/ORD 919.541.7828	WWW/Internet: Anticipated availability date April 15, 1996.	Selected data summaries and reports to be placed on WWW. Validated data available through OAR, Susan Womble, (202) 233-9057.	
	PerFluorocarbon [?] Tracer Database	Data related to indoor air modeling and exposure assessments; a comprehensive compilation of several survey studies of residential house volumes, air exchange rates, and infiltration rates.	Office of Pollution Prevention and Toxics	Pat Kennedy (OPPT) 202.260.3916		As yet, undergoing quality assurance procedures by U.S. EPA contractor.	
	Source Ranking Database	Data related to indoor air and exposure assessment.	Office of Air and Radiation	Christine Cinelli (OAR) 202.260.3913			
	Temporal Allocation Factors File (TAFF)	Contains national level default temporal allocation factors for use in developing temporally allocated emissions inventories.	Office of Research and Development, NRMRL/APPCD/ECPB	Charles Mann 919.541.4593 mann.chuck@ epamail.epa.gov.	No Internet; ASCII text file; Microsoft Access		
	Solvent Alternatives GuidE (SAGE)	Computer software system using a decision tree format to identify cleaning options for various surfaces.	Office of Research and Development, NRMRL/APPCD/ECPB	Charles H. Darvin 919.541.7633 E-mail: none	ENVIROENE HTTP://WASTENOT. INEL.GOV:80/ ENVIROENE/ TTN (919)541-5742		
	Coating Alternatives GuidE (CAGE)	Computer-based software program, to assist in reducing VOC and air toxic emissions from metal parts and product coating operations.	Office of Research and Development, NRMRL/APPCD/ECPB	Michael Kosusko 919.541.2734 mkosusko@ engineer.aeerl. epa.gov	No Internet		
	Adhesives Alternatives GuidE (AAGE)	Computer-based package on selection and use of adhesives and low-and-no-VOC substitutes for solvent-based adhesives.	Office of Research and Development, NRMRL/APPCD/ECPB	Chester Vogel 919.541.2827 cvogel@engineer. aeerl.epa.gov	No Internet		

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks	
	Biogenic Emissions Inventory System (BEIS-II) Land Use Database	Contains county land use data for the contiguous 48 states.	Office of Research and Development, NRMRL/APPCD/ECPB	Christopher D. Geron 919.541.4639 cdg@aesn02.aeerl.epa.gov.	TTN BBS.RTPNC.EPA.GOV		
	Global Technology (GloTech)	Contains information on cost, environmental releases, and other technology data which can be compared, summed and ranked.	Office of Research and Development, NRMRL/APPCD/APB	Lee Beck 919.541.0617 lbeck@engineer.aeerl.epa.gov	HTTP://WWW.EPA.GOV/DOCS/CRB/BROCHURE/GWCB/GLOTECH.HTM.	Software only. Beta test 96.	
Water	Ocean Data Evaluation System (ODES)	Effluent, water quality and biological monitoring data from all coastal sewage treatment plants that have or may submit 301(h) reports.	Office of Marine and Estuarine Protection		EPA 503/8 -90-007		
	(NFTDR) National Fish Tissue Data Repository	Collection and storage of fish and shellfish contaminants data, will be part of a larger EPA database and computer systemOcean Data Evaluation System (ODES).	Office of Water		EPA/823/B-93/003 Accession Number: 212076		
	Biological Data Management System (BIOS)	National biological data management system linked to STORETresults from data requirements survey of field biologists.					
	Exposure Analysis Modeling System (EXAMS-II)	An interactive modeling system allowing specification and storage of chemicals properties and ecosystem characteristics for evaluation of probable aquatic fate of synthetic organic chemicals.	Office of Research and Development	Model Coordinator 706.546.3549			
	National Compendium of Freshwater Fish and Water Temperature Data (FISHTEMP)	Historical fish distribution data with accompanying water temperature data from 1930-present for over 300 species of freshwater fish from 250,000 locations in the U.S. Provides a nationwide compendium of freshwater fish populations in relation to water temperature regimes.	Office of Research and Development	John Eaton 218.720.5557			
	Aquatic Toxicity Information Retrieval (AQUIRE)	Information on the toxic effects of chemicals to aquatic organisms and plants; toxicity test results and related testing information for any individual chemical from laboratory and field aquatic toxicity tests; acute, sublethal and bioconcentration effects are recorded for fresh water and marine organisms.	Office of Research and Development	Christine L. Russom 218.720.5709			
	Assessment Tools for the Evaluation of Risk (ASTER)	Assists Regional and State regulators in performing ecological risk assessments; an integration of the AQUIRE toxic effects database and the QSAR system, a structure activity based expert system.	Office of Research and Development	Christine L. Russom 218.720.5709			
	Quantitative Structure Activity Relationships System (QSAR)	Expert system providing information on physical-chemical properties, fate and effects of organic chemicals to the environment.	Office of Research and Development	Christine L. Russom 218.720.5709			
	Food and Gill Exchange of Toxic Substances (FGETS)	Databases and model predicting temporal dynamics of a fish's whole body composition (g chemical/(g live weight fish)) of non-ionic, non-metabolized, organic chemicals that are bioaccumulated from water only, or water and food.	Office of Research and Development, NERL, ERD-Athens	Frank Stancil 706.546.3197	stancil.frank@epamail.epa.gov		Database/Models

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	Worl Wide Web/ Internet / Source	Remarks	
	Lake Analysis Management System (LAMS)	A system of data bases and models, with GIS capability, developed for the Great Lakes and watersheds. Data bases and models, containing Great Lakes data collected since 1971 and Canadian Great Lakes data collected since 1968, assist in environmental decision-making for the Great Lakes.	Office of Research and Development	Russell G. Kreis 313.692.7615			Database/Models
	Green Cross Solar (GCSOLAR)	Computes direct photolysis rates and half-lives of pollutants in the aquatic environment.	Office of Research and Development	Model Coordinator 706.546.3549			Model
	Stream Quality Model (QUAL2E)	Permits simulation of water quality constituents in a branching stream system using a finite difference solution to the one-dimensional advective-dispersive mass transport and reaction equation; conceptual representation is a stream reach divided into a number of subreaches or computational elements equivalent to finite difference elements.	Office of Research and Development	Model Coordinator 706.546.3180			Model
	River Reach Files (RF1, RF2, RF3)	A series of hydrologic databases of the surface waters of the continental United States and Hawaii. The structure and content of these databases were created to establish hydrologic ordering, to perform hydrologic navigation for modeling applications, and to provide a unique identifier for each surface water feature, i.e., the reach code.	Office of Water	Karen Klima klima.karen@ epamail.epa.gov 800.424.9067	http://www.epa.gov/ OWOW/gis/reach. html		
	The Waterbody System (WBS)	Database provides convenient means for storing assessment information organized around water quality resource units called waterbodies. Assessment information is entered on beneficial use status or causes and sources of pollution for each waterbody. The program can generate lists and summary tables for preparation of Section 305(b) reports.	Office of Water	Karen Klima klima.karen@ epamail.epa.gov	http://www.epa.gov/ OWOW/NBSFlash/ NBSFlash.html		
	STOrage and RETrieval System for Water and Biological Monitoring Data (STORET)	Utility maintained for storage and retrieval of parametric data pertaining to the quality of waterways within and contiguous to the US. Has evolved to a comprehensive family of systems, performing range of functions, including data availability summaries; tabular data reports; statistical data analyses; graphics and maps; and data preparation for down-load to other systems.	Office of Water	Phil Lindenstruth	http://www.epa. gov/OWOW/ STORET/ STORET@ epamail.epa.gov		
	*National Estuaries Program (NEP)	Identifies nationally-significant estuaries to protect and improve their water quality and to enhance their living resources [21].					
	*National Water Quality Monitoring Program	US water quality monitoring and assessment guidelines, protocols and programs for Clean Water Act 305(b) reporting.					
	Discharge Monitoring Report-Quality Assurance Studies (DMR-QA)	Annual analysis of inorganic synthetic wastewater samples by 7,000 to 7,500 major National Pollution Discharge Elimination System (NPDES) dischargers; provides basis for evaluating quality of routine monitoring required in NPDES permits.	Office of Research and Development	Paul Britton 513.569.7216			Database
	Water Pollution Laboratory Performance Evaluation Studies (WP)	Semi-annual analyses of synthetic wastewater samples for up to 80 organic and inorganic analytes by routine wastewater analysis laboratories.	Office of Research and Development	Paul Britton 513.569.7216			

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks
	Water Supply Laboratory Performance Evaluation Studies (WS)	Semi-annual analyses of synthetic drinking water samples for up to 170 organic and inorganic analytes by laboratories seeking U.S. EPA certification to analyze drinking waters.	Office of Research and Development	Paul Britton 513.569.7216		
	Treatability Database	Compendium of peer-reviewed treatability data on a wide range of environmental contaminants, largely in an aqueous medium but with some information on hazardous wastes.	Office of Research and Development, NRMRL/STD	Jerry Waterman 513.569.7834		
	North American Wetlands for Water Quality Treatment	Compilation of information of wetlands used to treat wastewater.	Office of Research and Development, NRMRL/WSWRD/ WQMB	Don Brown 513.569.7630 brown.donald@ epamail.epa.gov		
	EARNS	Notifications of oil and hazardous substance releases to the environment.				
	*HUC Code Digitization	Details USGS HUCs (watershed boundaries) to sub-watershed delineations.				
	PCS	Data on National Pollutant Discharge Elimination System permit-holding facilities.				
Soil	Dougherty Plain Database	Four-year database on leaching of aldicarb, metolachlor and bromide in a 3.9 ha peanut field. Data includes application rates, pesticide/bromide concentration profiles, plant uptake, soil water retention, hydraulic conductivity and daily weather.	Office of Research and Development (ORD), NERL, ERD-Athens	Charllie Smith 706.546.2247 smith.charllie@ epamail.epa.gov		
	Watkinsville Database	Four-year database on runoff of atrazine, cyanazine, diphenamid, propazine, trofluralin, 2,4-D and fertilizer from four small agricultural watersheds (1.32.7 ha) for use in development and testing of runoff transport models. Data include pesticide and fertilizer runoff by rainfall event, application rates, soil characteristics, pesticide and fertilizer concentration profiles, and daily weather records.	Office of Research and Development (ORD), NERL, ERD-Athens	Charllie Smith 706.546.2247 smith.charlie@ epamail.epa.gov		
	Unsaturated Soil Hydraulic Database (UNSODA)	Database provides field and laboratory measured soil hydraulic and physical properties for use in unsaturated zone models. The database includes the Retention Curve (RETC) code for the estimation of hydraulic parameters for use in these models.	Office of Research and Development , NRMRL/SPRD Center for Subsurface Modeling Support (CSMoS).	Joe. R. Williams 405.436.8608	http://www.epa.gov/ ada/kerrlab.html	Database was developed in cooperation with the USDA Soil Salinity Laboratory in Riverside, California. The database is currently being made ready for distribution via the CSMoS WWW Site listed.
Soil and Ground Water	Model Annotation and Retrieval System (MARS)	Database of subsurface computer models containing information on over 500 models for the unsaturated and saturated zone. System allows the user to search for models by specific model application needs.	Office of Research and Development, NRMRL/SPRD Center for Subsurface Modeling Support (CSMoS).	Joe. R. Williams 405.436.8608	Currently not available on the WWW site.	This database was developed under cooperative agreement with the International Ground Water Modeling Center (IGWMC) and is being maintained and marketed by IGWMC. The updated version can be obtained by contacting IGWMC at (303) 273-3103.

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks
Multi-Media	*Environmental Monitoring and Assessment Program (EMAP)	Interdisciplinary, multiple US ecological resource program to monitor status and trends with known confidence. Embodies a comprehensive information management system.	National Health and Environmental Effects Research Laboratory	Gilman Veith, NHEERL Tom Murphy, NHEERL Robert Shepanek, NCEA	http://epawww.epa.gov/emaphome	
	Agricultural Lands	Status and trends in the condition of the nation's agricultural lands which include land used for crops, pasture, and livestock; adjacent land; and the associated atmosphere, underlying soils, ground water, and drainage networks.	National Health and Environmental Effects Research Laboratory	Charles Lee Campbell 919.515.3311	http://www.epa.gov/emap/html/resrcgrp/agroland/	
	Estuaries	Status and trends in the condition of the nation's estuaries extending inland to the head of tide. This includes to coastal embayments, bays, inland waterways, tidal rivers, coastal wetland areas, and salt-water marshes.	National Health and Environmental Effects Research Laboratory	Kevin Summers 904.934.9244	http://www.epa.gov/emap/html/resrcgrp/estuary/ http://dolphin.gbr.epa.gov	
	Forests	Status and trends in the condition of the nation's forests, including areas formerly covered by trees, but not currently built-up or developed for agricultural use.	National Health and Environmental Effects Research Laboratory	Sam Alexander 919.549.4020 Beth Eastman 919.549.4059	http://www.epa.gov/-emap/html/resrcgrp/f-orests/	
	Great Lakes	Status and trends in the condition of the Great Lakes, the nation's largest and most unique group of surface waters.	National Health and Environmental Effects Research Laboratory	Stephen Lozano 218.720.5594	http://www.epa.gov/-emap/html/resrcgrp/-greatlak/	
	Landscape Ecology	Studies of the structure, function, pattern, and changes in the nation's heterogenous land areas using remote sensing and geographic information systems.	National Health and Environmental Effects Research Laboratory	Bruce Jones 702.798.2671	http://www.epa.gov/-emap/html/resrcgrp/l-andecol/	
	Rangelands	Status and trends in the condition of the nation's deserts, grasslands, and other arid lands.	National Health and Environmental Effects Research Laboratory	Dan Heggem 702.798.2278	http://www.epa.gov/-emap/html/resrcgrp/r-angland/	
	Surface Waters	Status and trends in the condition of the nation's surface waters including lakes, streams and rivers downstream to the head of tide, and inland wetland areas.	National Health and Environmental Effects Research Laboratory	Steve Paulsen 503.754.4428	http://www.epa.gov/emap/html/resrcgrp/surfwatr/	
	Interagency Taxonomic Information System (ITIS)	Provide an electronic, taxonomic reference that promotes scientific excellence and is fully supported by the world taxonomic community.	National Health and Environmental Effects Research Laboratory	Gary Collins 513.569.7174	http://trident.ftc.nrcs.usda.gov:80/itis/dev/	
	Methods	Database on selection and development of appropriate methods for collecting data. Encourages the use of standardized methods within resource groups to ensure comparability of data and helps to develop strategies for standardizing methods across resource groups.	National Health and Environmental Effects Research Laboratory	Gary Collins 513.569.7174	http://www.epa.gov/docs/emap/html/coordgrp/methods/	
	Indicators	Database on basic research concerning the development and selection of biological and other indicators for use in EMAP.	National Health and Environmental Effects Research Laboratory	Kay Austin	http://www.epa.gov/docs/emap/html/coordgrp/indicatr/	
	*Multi-resolution Land Characteristics Consortium (MRLC)	Interagency program to produce a consistent US land cover database; integrates data and information across temporal and spatial scales.	National Health and Environmental Effects Research Laboratory	Denice Shaw 919.541.2698	http://www.epa.gov/docs/grd/mrlc/	

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks	
	Exposure Models Library and Integrated Model Evaluation System (IMES)	Selection of over 70 fate and transport models for exposure assessments in various environmental media. Facilitates selection of best fate model for a particular risk assessment problem as well as information on validation of the models.	Office of Research and Development, NCEA-Washington.	Richard Walentowicz 202.260.8922	EPA/600/C-92/200		
	*Developmental and Reproductive Toxicity (DART) Database	On-line, bibliographic information and data on developmental and reproductive toxicity.	Office of Research and Development, NCEA-Washington	Carol Kimmel 202.260.7331			
	ORD Bibliography	Searchable, downloadable, electronic database of ORD publication citations with abstracts (1976-present).	Office of Research and Development, NRMRL/TTSD	Randy Revetta 513.569.7358		Accessible through ORD Bulletin Board (ORDBBS). (513) 569-7610 (513) 569-7700 (513) 569-7272 (voice)	
	*Ecotoxicology Database Retrieval System (ECOTOX)	Computer-based system that provides chemical-specific toxicity values for aquatic life, terrestrial plants, and wildlife; useful in developing consistent ecosystem management decisions within the EPA and other federal, state, local, and international governmental agencies.	Office of Research and Development, NHEERL	Christine L. Russom 218.720.5709 Steve Bradbury 218.720.5527	http://www.epa.gov/docs/ORD/BBS.html	Developed by ORD with funding from the U.S. Department of Defenses Strategic Environmental Research Defense Program.	
	EMMI	Database of ~ 2,600 EPA regulated substances; includes regulatory limits, method detection limits and abstracts.					
	Soil Transport and Fate Database and Model Management System (STF)	Database providing degradation, transformation, toxicity, bioaccumulation, and partitioning information on approximately 400 chemicals (organic and some inorganic) in the soil environment. Includes the RITZ and VIP models.	Office of Research and Development, NRMRL, Center for Subsurface Modelling Support (CSMoS)	David S. Burden 405.436.8606	http://www.epa.gov/ada/kerrlab.html		Database/Models
	Pesticide Assessment Tool for Rating Investigation of Transport (PATRIOT)	Datasets on daily rainfall (10 years) from 200 NOAA weather stations, properties and occurrences of 8000 soils, pesticides properties and cropping practices complement the PATRIOT Model for rapid assessment of groundwater vulnerability to pesticide contamination in the contiguous U.S.	Office of Research and Development, NERL, ERD-Athens	Frank Stancil 706.546.3130 stancil.frank@epamail.epa.gov			Databases/Model
	Pesticide and Industrial Chemical Risk Analysis and Hazard Assessment (PIRANHA) Program	Includes a geographical database for locating biological resources potentially at risk from pesticides; useful in applying chemical, ecotoxicological and environmental sciences to ecological risk assessment.	Office of Research and Development, NERL, ERD-Athens	Lawrence Burns 706.546.3511 burns.lawrence@epamail.epa.gov			Databases/Model
	National Environmental Supercomputing Center (NESC)	Computational resources necessary to carry out scientific modeling efforts; provides computer support for many critical environmental models.	Director, National Environmental Supercomputing Center	517.894.7600 Fax: 894.7676 Cullati.Art@epamail.epa.gov	http://www.epa.gov/accessepa/chapter8/s1-1.html		
	Sparc Provides Automated Reasoning in Chemistry (SPARC)	Expert system for estimating chemical and physical reactivity estimates on a broad scope and inexpensively.	Office of Research and Development (ORD), NERL, ERD-Athens	Mac Long 706.546.3349 long.mac@epamail.epa.gov			Expert System
	Environmental Fate Constants Database (FATE)	Provides interactive retrieval of kinetics and equilibrium constants used in modeling and assessing chemical (300 and growing) fate in the environment.	Office of Research and Development (ORD), NERL, ERD-Athens	Brenda Kitchens 706.546.3198 kitchens.brenda@epamail.epa.gov			

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks
	Guide to National Environmental Statistics	Tool for locating sources of national-level environmental statistics available from the U.S. Government; a reference to national-level, time-series environmental statistics; not intended to be inclusive; data sources from 23 offices in 7 agencies: USDA, DOI, DOC, DOT, DOE, EPA, DHHS.	Office of Policy, Planning, and Evaluation	Brand Niemann 202.260.3726		
	Integrated Risk Information System (IRIS)	EPA's primary vehicle for communication of chronic non-caner and cancer health hazard information for over 500 substances. IRIS contains substance specific hazard identification and dose-response information.	Office of Research andDevelopment, NCEA-CIN	513.569.7254 Hotline # FAX 513.569.7159	IRIS Hotline 513.569.7254	
	Total Human Exposure Risk database and Advanced Simulation Environment (THERdbASE)	Data and model management system containing such total human exposure information as census data, state activity pattern studies, EPAs National Human Activity Pattern Study, TEAM and NOPES data, inhalation and dermal models, or indoor air models.	Office of Research and Development (ORD), NERL CRD	John Quackenbos 702.798.2442	http://eeyore.lv-hrc.nevada.edu/therdbase.html	
	Quantitative-Structure Toxicity Relationships (QSTR)	Enables computation of a probable value of toxicity for a given chemical structure, for various health endpoints.	Office of Research andDevelopment, NCEA-CIN	R. Bruce 513.569.7569		
	Treatability Data Base (TDB)	Assist users in selecting technologies or estimating the degree of treatment achievable for specific chemicals in all types of waters, wastewaters, soils, sediments, and debris; data extracted from a variety of sources.	Office of Research andDevelopment	Glenn M. Shaul 513.569.7408		
	Government Information Locator Service (GILS)	Virtual card catalog" of government information. The EPA's GILS website offers access to records that describe the agency's information resources.			http://www.epa.gov/gils/	
	ACCESS EPA	Directory provides contact information and description of services for more than 300 of EPA's major information resources, including databases, models and EPA libraries.	Office of Administration and Resources Management		http://www.epa.gov/docs/Contacts/Access/chapter5.txt.html; EPA 220-B-93-008	
Hazardous Waste and Superfund	CERCLIS	Superfund database with information on all identified U.S. hazardous waste sites.				
	Sediment Toxicity Database	Sediment chemistry and toxicity measurements from saltwater and freshwater sites.	Office of Research and Development, NECA-Washington.	Sue Norton 202.260.8922		
	OHMTADS	Chemical properties, regulatory information, toxicity and safety data, and response information on hazardous substances.				
	RODS	Tracks site cleanups under the Superfund program.				
	RCRIS	Tracks events and activities related to facilities that generate, transport, treat, store, or dispose of, hazardous waste.				
	PIN	Pesticide monitoring inventory, environmental fate, and effects data.				
	TRI	Facility and substance identification for toxic chemicals released directly to air, water or land, or that is transported off-site.				

Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks	
	Bioremediation in the Field Search System (BFSS)	Information on over 500 sites where bioremediation is being tested, implemented, or has been completed.	Office of Research and Development, NRMRL/TTSD	Fran Kremer 513.569.7346		Accessible through Alternative Treatment Technology Information Center (ATTIC); (703) 908-2137. Also accessible through ORD Bulletin Board (ORDBBS);(513) 569-7610; (513) 569-7700 (513) 569-7272 (voice).	
	Treatment Technology Database	Provides information on biological, chemical, physical, solidification/stabilization, and thermal treatment processes.	Office of Research and Development, NRMRL/WSWRD	Daniel Sullivan 908.321.6677		Accessible through Alternative Treatment Technology Information Center (ATTIC).Hotline (703) 908-2137	
	Treatability Study Database	Compendium of peer-reviewed treatability data on a wide range of environmental contaminants, largely in an aqueous medium, but with some information on hazardous wastes.	Office of Research and Development, NRMRL/STD	Jerry Waterman 513.569.7834		Accessible through Alternative Treatment Technology Information Center (ATTIC).Hotline (703) 908-2137	
	Underground Storage Tank Database	Information on underground storage tank corrective actions, surface spills, and remedial action. Allows retrieval of information to help select appropriate, cost-effective solutions.	Office of Research and Development, NRMRL/WSWRD/ UWMB	Daniel Sullivan 908.321.6677		Accessible through Alternative Treatment Technology Information Center (ATTIC).Hotline (703) 908-2137	
	Oil/Chemical Spill Database (SPILLS)	Information on the treatment and disposal of spilled oil and chemicals.	Office of Research and Development, NRMRL/WSWRD/UW-MB	Daniel Sullivan 908.321.6677		Accessible through Alternative Treatment Technology Information Center (ATTIC).Hotline 703.908.2137	Database/Model
	Alternative Treatment Technology Information Center (ATTIC)	Provides information on innovative treatment technologies. Includes case studies providing information on such technologies in support of hazardous waste clean-up.	Office of Research and Development (ORD), NRMRL-Edison.	Daniel Sullivan 908.321.6677	http://www.epa.gov/ attic/attic.html		
	Vendor Information System for Innovative Treatment Technologies (VISITT)	Current information on availability, performance and cost of innovative technologies to remediate soil and groundwater contaminated by hazardous and petroleum wastes.	Office of Solid Waste and Emergency Response, Technology Information Office		Available within CLU-IN BBS (301.589.8366; File=visit3.zip)		
Other	Research Tracking System (RTS)	Tracks research technical information: projects, tasksmilestones; publications produced by Mid-continent Ecology Division scientists.	Office of Research and Development	Judy L. Stagner 218.720.5605			
	Criteria Reference Information Bank (CRIB)	Contains bibliographic information identifying sources cited in air quality criteria and other documents produced by the National Center for Environmental Assessment - RTP Office (NCEA-RTP Office).	Office of Research and Development	Douglas B. Fennel 919.541.3789			
	Athens Chemical Inventory System (ACIS)	In-house chemical inventory is maintained for the purpose of chemical hygiene and environmental compliance.	Office of Research and Development	Frank Stancil 706.546.3197			
	Technical Assistance Database (TAD)	Internal system used to maintain a database of technical support provided to the user community, inside and outside of EPA, by the Center for Exposure Assessment Modeling (CEAM).	Office of Research and Development	Robert Ryans 706.546.3306			
	NIST/EPA/NIH/Mass Spectral Database	Library of mass spectra used on-line by mass spectrometers for compound identification.	Office of Research and Development, NERL CRD/ASB/ACR	Ed Heitmar 702.798.2626	N/A	Database is copyrighted by other organizations and licensed to ORD only for use with specific instruments.	

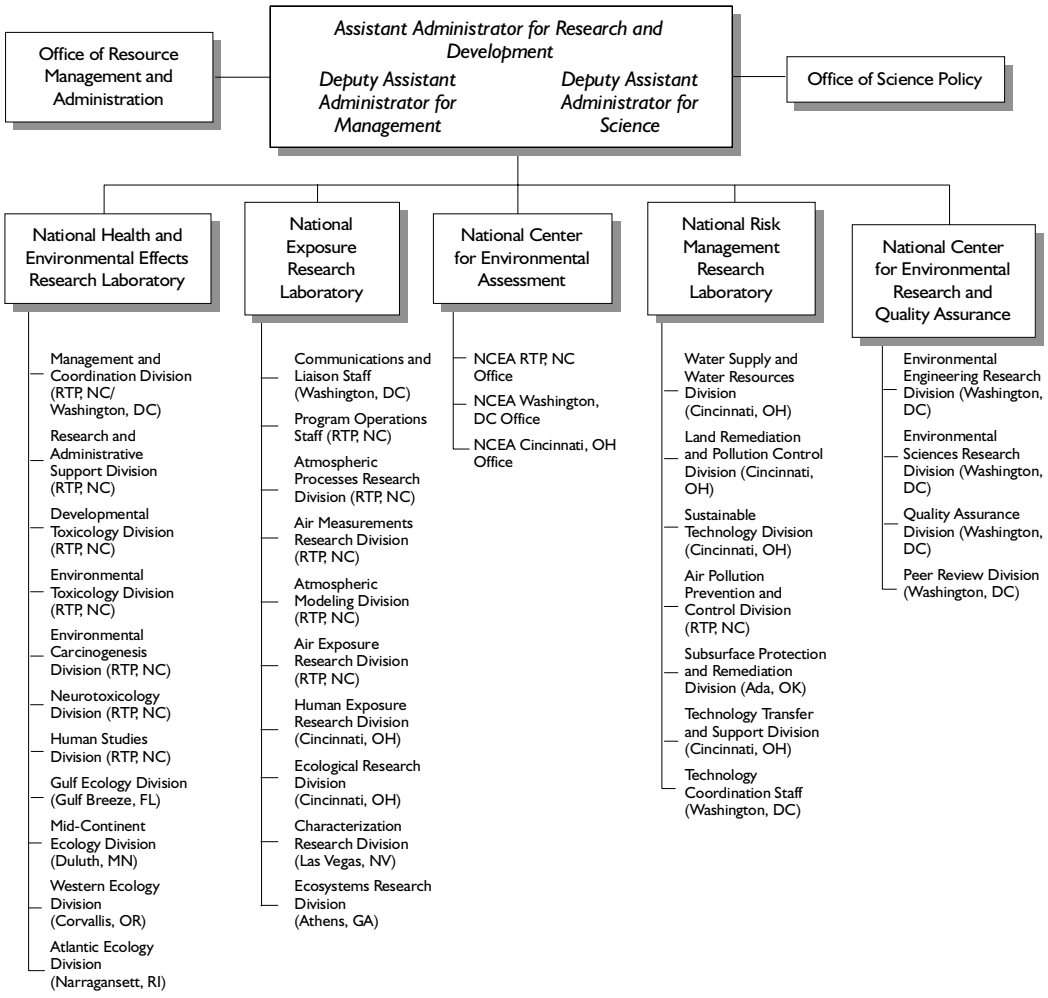
Medium	Name (*denotes Multiagency)	Description [Number of Sites]	Responsible EPA Organization	Contact/ Telephone	World Wide Web/ Internet/Source	Remarks
	Report Locator Data Base	Identifies aerial photographic interpretation projects and reports completed by Environmental Photographic Interpretation Center (EPIC) over the past 20 years.	Office of Research and Development, NERL, CRD			For internal use only. Database currently on VAX; but needs to be converted for use under ORACLE software.
	Pesticide Treatability Database	Compendium of information specific to pesticides, their formulation, and treatment options.	Office of Research and Development, NRMRL/STD/MTB	Dave Ferguson 513.569.7518		
	Enviroene	Hosts an expert architecture known as the Solvent Umbrella, allowing users to access solvent alternative information through a single, easy-to-use command structure.	EPA/DOD/DOE	Idaho National Engineering Laboratory	http://www.epa.gov/I-daho.html	

Please send any comments, additions, modifications or corrections to Sidney Draggan via electronic mail.

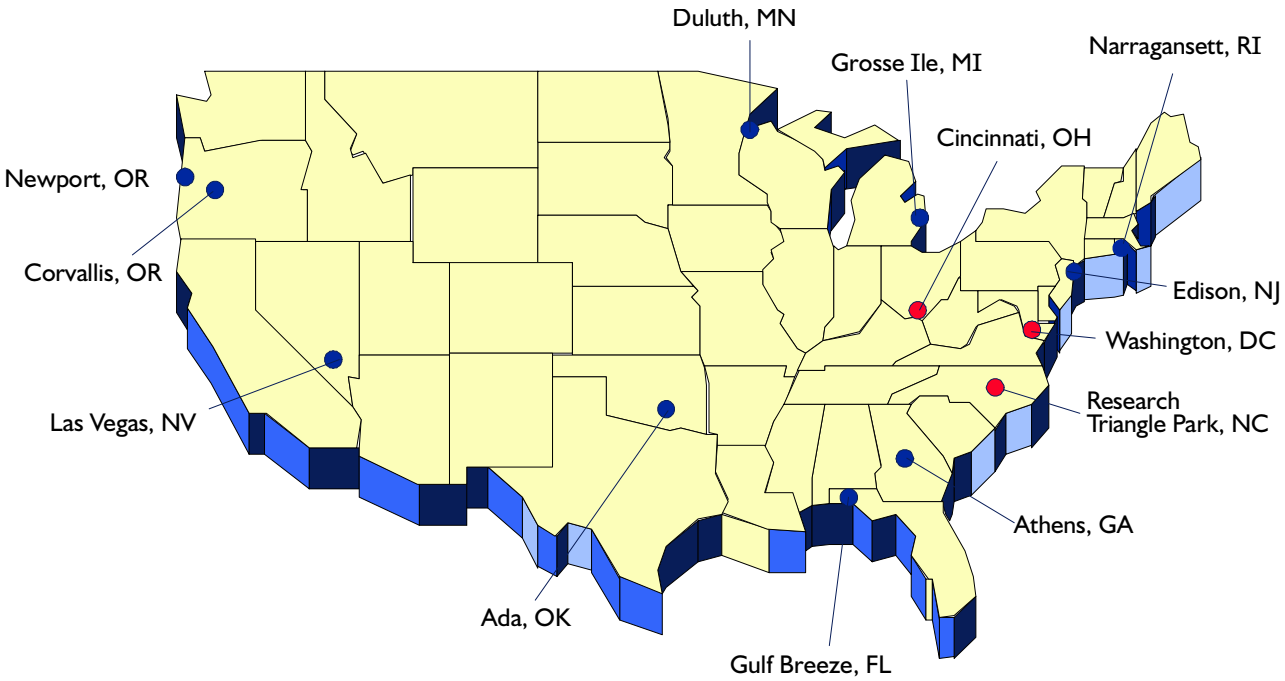
Appendix D:
The Office of Research and Development Organization

ORD's organization, depicted below, mirrors the risk assessment/risk management paradigm. The locations of ORD's National Laboratories, Centers, and Offices are shown on the following page; their functions are described in the ORD Strategic Plan.

ORD's New Risk-Based Organization



Location of ORD's National Laboratories and Centers



This document was developed by the ***ORD Information Management Strategy Group***, composed of:

Laurie Schuda (Chair), ORD Immediate Office of the Assistant Administrator
Owen Bricker, ORD Environmental Monitoring and Assessment Program
Sidney Draggan, EPA Office of the Administrator
Carol Finch, ORD Immediate Office of the Assistant Administrator
Dick Garnas, ORD National Exposure Research Laboratory
Stephen Hale, ORD National Health and Environmental Effects Research Laboratory
John Ireland, ORD National Risk Management Research Laboratory
Linda Kirkland, ORD National Center for Environmental Research and Quality Assurance
Scott Minamyer, ORD National Risk Management Research Laboratory
Bob Shepanek, ORD National Center for Environmental Assessment
Charissa Smith, ORD Office of Resources Management and Administration
Linda Tuxen, ORD National Center for Environmental Assessment
Jeanette Wiltse, ORD National Center for Environmental Assessment
Steve Young, EPA Office of Information Resources Management

United States Environmental Protection Agency
(8101)
Washington, DC 20460

Official Business
Penalty for Private Use
$300